UNIVERSITY OF CALIFORNIA
Santa Barbara

# First-principles investigations of III-nitride bulk and surface properties

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Materials

by

Cyrus Eduard Dreyer

Committee in Charge:

Professor Chris G. Van de Walle, Chair

Professor James S. Speck

Professor Steven P. DenBaars

Professor Umesh K. Mishra

June 2014

The Dissertation of
Cyrus Eduard Dreyer is approved:

---

Professor James S. Speck

---

Professor Steven P. DenBaars

---

Professor Umesh K. Mishra

---

Professor Chris G. Van de Walle, Committee Chairperson

June 2014

First-principles investigations of III-nitride

bulk and surface properties



by

Cyrus Eduard Dreyer

*To JC, through thick and thin.*

# Acknowledgements

I would like to express my gratitude to the many people who made this work possible. First and foremost is my advisor, Prof. Chris Van de Walle. I have learned so much from Chris about materials science, physics, and how to be a better scientist. He taught me to be careful and skeptical about what I read and hear, and especially about my own results. He pushed me to become immersed in my work when I was lacking direction. I would not be at the point I am today without Chris to direct me, and, as I move forward, I will take the lessons he taught me along to my future endeavors.

Also, I would like to thank Dr. Anderson Janotti for all of his guidance. Throughout the PhD process, Anderson provided me with everything from practical advice about running calculations, to interesting discussion about half-baked ideas, to psychological support and encouragement during times of frustration. He was always available to answer questions large or small and proofread my abstracts and manuscripts.

I am indebted to all of the postdocs that I had a chance to interact with in Chris' group. Their wide range of expertise and constant willingness to answer questions and provide guidance was invaluable. Near the beginning of my PhD, my incessant questions about calculations and DFT theory were answered patiently by Drs. Maosheng Miao, Patrick Rinke, Manos Kioupakis, Poul Moses, Christian Carbogno, Lars Ismer, and Khang Hoang. Dr. Daniel Steiauf suffered through innumerable troubleshooting

sessions and helped navigate me through the VASP source code, as well as answering my basic physics questions. Later in my time at UCSB, I had many interesting discussions with Dr. Hartwin Peelaers, and received a lot of guidance on DFT, physics, and defects in semiconductors from Dr. Audrius Alkauskas. Dr. Anyindia Roy was instrumental in helping me develop knowledge of the modern theory of polarization. Finally I acknowledge fruitful conversations with Drs. Hazem Abu Farsakh and Minseok Choi.

In addition to the postdocs, I learned an enormous amount from the current and former students in Chris' group. Early on, Drs. Justin Weber and Joel Varley gave me a substantial amount of advice and guidance. Throughout my PhD work, Drs. Qimin Yan and Jack Lyons taught me about properties of nitrides, strain, and point defects. I have had fruitful discussion and collaborations that have shaped my research and education with the current students in the group, Luke Gordon, Lars Bjaalie, Karthik Krishnaswamy, Wennie Wang, Hiral Taylor, Jimmy Shen, and Patrick McBride.

Also, my PhD would not have been possible without the support of the funding agencies. I would like to thank the Department of Energy and the National Science Foundation for the financial support that funded my research, as well as the Solid State Lighting and Energy Center for the support and for exposing me to the semiconductor and lighting industry.

I would also like to thank Profs. Jim Speck, Umesh Mishra, and Steve DenBaars for serving on my committee, and for their help and feedback during the dissertation

process, as well as Profs. Alexey Romanov and David Vanderbilt for their feedback on my research.

Finally, my work at UCSB also would not have been possible without the support of my family and friends. In particular my parents Erich and Feroza Dreyer as well as my sister Naomi. And, of course, my girlfriend Jen Cano who has been a constant source of support and inspiration through the ups and downs of grad school; I could not have done it without her.

# Curriculum Vitæ

## Cyrus Eduard Dreyer

### Education

| | |
|---|---|
| 2014 | Ph.D. in Materials, University of California, Santa Barbara. |
| 2009 | B.S., High Honors, in Engineering Science, Physics, and Math, University of Virginia. |

### Awards

| | |
|---|---|
| 2014 | Selected to delegation attending California Graduate Research Advocacy Day |
| 2013 | Solid State Lighting and Energy Center Outstanding Graduate Student Research Achievement Award |

### Publications

1. Cyrus E. Dreyer, Willaim V. Chiu, Robert H. Wagoner, and Sean R. Agnew, "Formability of a More Randomly Textured Magnesium Alloy Sheet: Application of an Improved Warm Sheet Formability Test," *Journal of Materials Processing Technology* **210**, 37 (2010).
2. Cyrus E. Dreyer, Anderson Janotti, and Chris G. Van de Walle, "Effects of strain on the electron effective mass in GaN and AlN," *Applied Physics Letters* **102**, 142105 (2013).
3. Cyrus E. Dreyer, Anderson Janotti, and Chris G. Van de Walle, "Absolute surface energies of polar and nonpolar planes of wurtzite GaN," *Physical Review B* **89**, 081305(R) (2014).
4. Cyrus E. Dreyer, Anderson Janotti, and Chris G. Van de Walle, "Band alignments and polarization properties of BN polymorphs," *Applied Physics Express* **7**, 031001 (2014).
5. Karthik Krishnaswamy, Cyrus E. Dreyer, Anderson Janotti, and Chris G. Van de Walle, "Structure and energetics of $LaAlO_3$ (001) surfaces," *Submitted.*

### Press Coverage

1. "Aiding nitride growth with accurate surface energies," *Compound Semiconductor*, May 19, 2014.

**Contributed talks**

| | |
|---|---|
| 2010 | The Minerals, Metals, and Materials Society Meeting, Seattle, WA<br>"Extracting post-uniform constitutive behavior from high temperature tensile test data" |
| 2011 | American Physical Society March Meeting, Dallas, TX<br>"Effects of strain on effective masses in GaN and AlN" |
| 2011 | The Minerals, Metals, and Materials Society Electronic Materials Conference, Santa Barbara, CA<br>"Effects of strain on effective masses in GaN and AlN" |
| 2012 | American Physical Society March Meeting, Boston, MA<br>"Surface energies and cracking in GaN" |
| 2012 | Annual Meeting of the California-Nevada Section of the APS, San Luis Obispo, CA<br>"Absolute surface energies of polar and non-polar planes in GaN" |
| 2012 | Nitrides Seminar, Santa Barbara, CA<br>"Absolute surface energies of polar and non-polar planes in GaN" |
| 2013 | American Physical Society March Meeting, Baltimore, MD<br>"Absolute surface energies of polar and non-polar planes in GaN" |
| 2013 | International Conference on Nitride Semiconductors, Washington D.C.<br>"Absolute surface energies of polar and non-polar planes in GaN" |
| 2013 | Solid State Lighting and Energy Center Review, Santa Barbara, CA<br>"Absolute surface energies, fracture toughness, and cracking in nitrides" |
| 2014 | American Physical Society March Meeting, Denver, CO<br>"Absolute surface energies, fracture toughness, and cracking in nitrides" |
| 2014 | Nitrides Seminar, Santa Barbara, CA<br>"Spontaneous polarization in III-nitrides and the importance of a consistent reference structure" |

# Abstract

## First-principles investigations of III-nitride bulk and surface properties

Cyrus Eduard Dreyer

The III-nitride semiconductors, including AlN, InN, GaN, and BN have been demonstrated as technologically exciting materials for a wide range of device applications. With band gaps that span the visible range, GaN, InN, and InGaN alloys are used for high efficiency light emitting diodes for general lighting, as well as laser diodes for optical storage. The wide gaps, large band offsets, and polarization fields in AlN, GaN, and AlGaN alloys are promising for high-frequency, high-power transistors with applications in power conversion and radio frequency amplifiers.

Despite the plethora of attractive material parameters of the III-nitride materials there are several issues that significantly limit the efficiency of devices and range of possible applications. In this study, we use first-principles electronic structure calculations to explore several of these properties relevant to understanding growth, processing, and device design.

Arguably the most detrimental issue in this material system is the lack of widely available, cost-effective substrates for the growth of films and devices. Heteroepitaxy, as well as the lattice mismatch between the layers of different III-nitride alloys in het-

erostructures, results in residual stresses in films and devices. Such stress will alter the electronic structure of the materials, so it is necessary for device design to be able to quantify these effects. We explore the influence of strain on the effective mass of carriers in GaN and AlN, a parameter that is tied to the conductivity. In addition, films under tensile strain can crack if the strain energy is sufficient. We explore the propensity for AlN, GaN, and AlGaN to crack on different crystallographic planes.

There has been significant work done to overcome the issue of residual strains in III-nitride films, both through the growth of bulk crystals for substrates, and the growth of structures such as nanowires that avoid many of the thickness and alloy-content limitations of epitaxial layers. Also, selective area growth followed by lateral overgrowth techniques is used to produce high-quality GaN for laser applications. In all of these cases, growth involves multiple growth planes, and the resulting morphologies will be determined by the relative growth rates of the planes present. One of the fundamental parameters that determines the stability and growth rate on a given plane is the absolute surface energies of that plane. Calculations of such surface energies are extremely challenging; we have succeeded in obtaining such values, for the first time, for the most technologically important planes in GaN.

Finally, residual stresses contribute to a piezoelectric polarization moment in the III-nitrides in the most common growth direction, and even in the absence of such strain, the wurtzite crystal structure has a spontaneous polarization moment. Polarization dif-

ferences between layers of heterostructures allow for strong carrier confinement and the formation of two dimensional electron gas (2DEG) with high densities at AlGaN/GaN interfaces, exploited in high electron mobility transistors (HEMTs). The effect of polarization can also be detrimental, for example causing the quantum confined Stark effect in quantum wells of light emitting diodes (LEDs). In both HEMT and LED devices, accurate values of the magnitudes of the spontaneous and piezoelectric polarization constants are required for a fundamental understanding and design of III-nitride devices. We have discovered that the set of spontaneous polarization coefficients that has been widely used for device design and analysis suffers from inconsistencies in the choice of reference structure, and have derived a new reliable and accurate set of constants to describe the polarization properties of III-nitrides. We discuss the results in the context of experimental observations.

---

Professor Chris G. Van de Walle
Dissertation Committee Chair

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 The III-nitrides: A new semiconductor revolution

A semiconductor is a material with a band gap, in which the conductivity can be tuned by doping, gating, temperature, strain, or other methods. Semiconductors are the primary building block of electronic devices including transistors, light emitters, solar cells, and sensors. By far, the semiconductor that has been most important in the development of electronic devices has been silicon, which is the material that makes up the quintillions of transistors that power modern computation. The success of Si has mainly been a result of exquisite control of defects and processing allowing the growth of

near-perfect crystals, the tuning of the electronic structure through strain engineering, and the efficacy of doping to make highly conductive layers.

However, there are several aspects of Si that limit its possible applications. The band gap is indirect, which precludes it as a possible material for optoelectronics because of its poor radiative recombination efficiency. Also, the band gap is around 1.1 eV,[1] which limits the fields that it can sustain and still be insulating, which is an issue for high-power, high-frequency transistor applications.

To overcome these material limitations, there has been significant work on compound semiconductors, traditionally involve elements from group-III and group-V (III-V) of the periodic table (GaP, InP, GaAs, GaSb, and so on) or group II and group IV (II-IV) (ZnO, CdTe, ZnS, *etc.*)[1] These materials have a diverse set of properties including indirect and direct band gaps of various sizes.

Of all of the possible combinations, compounds with a group III metal (B, Al, Ga, or In) and nitrogen have been demonstrated as outstanding materials with technologically exciting properties. The III-nitrides have seen widespread commercial success; in contrast to Si and many other compound semiconductors (notably GaAs), this success has been in spite of highly defective material and relatively poor understanding of the fundamental materials properties. This makes the III-nitrides a distinctly different semiconductor revolution, in which understanding has lagged behind rapid commercialization. Also, this makes the III-nitrides a system with many interesting scientific

questions whose answers have significant impact to what is already a multi-billion-dollar industry.

### 1.1.1 Basic properties of III-nitrides

AlN, GaN, and InN have the wurtzite (wz) (space group $P6_3mc$) structure in their ground state.[1] The ground state of BN is a hexagonal (h) layered material, similar to graphite (space group $F43m$), though it can exhibit the zincblende (zb) ($F\bar{4}3m$) phase at high temperatures.[2] Also, because of the interest in BN alloyed with GaN or AlN,[3] the properties of wz-BN are important to determine how dilute alloys will behave.

A schematic of the wz structure is shown in Fig. 1.1. Each unit cell has two nitrogen atoms and two cations. The $u$ parameter indicated in Fig. 1.1 is an internal parameter, meaning that it can take on different values without altering the unit cell or point symmetry. We see that wz is hexagonal in the $c$ plane and is polar (alternating layers of cations and ions) perpendicular to the $c$ plane with "$ABABAB$...." stacking. When III-nitride device heterostructures are grown, the most common growth direction is the $c$ direction.

Most traditional semiconductor materials are cubic (Si and Ge have the diamond structure and GaAs, InP, *etc.* are zb); the difference in structure has significant consequences for the properties and applications. The reduced symmetry in the $c$ direction compared to cubic structures results in polarization moments in that direction that can

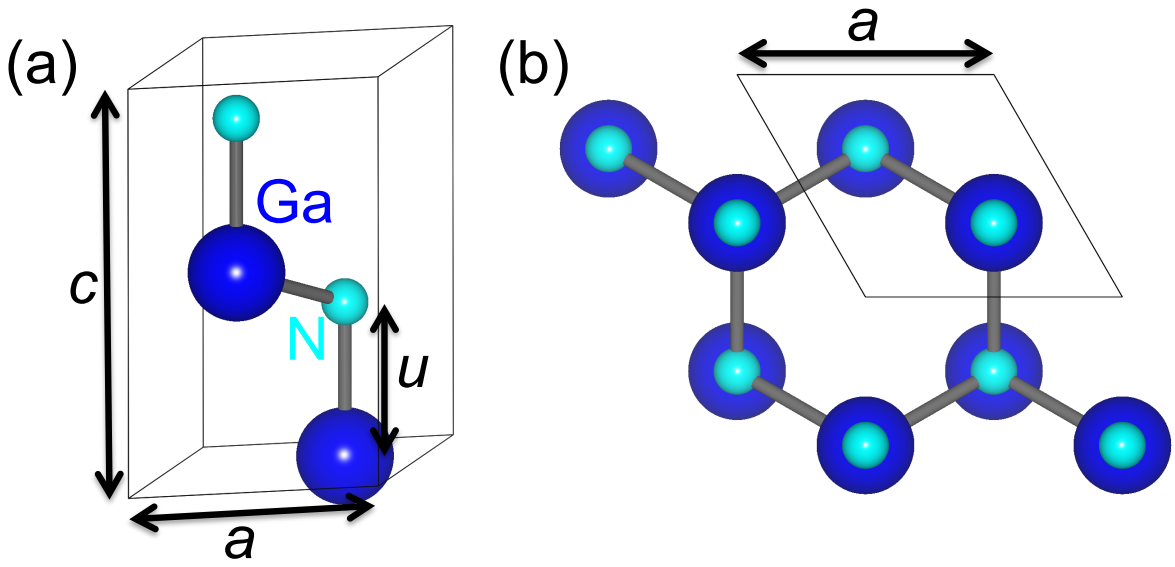


**Figure 1.1:** Ball-and-stick model of GaN in the wurtzite structure. (a) Side view illustrating structural parameters and (b) top view illustrating hexagonal structure in the $c$ plane.

be beneficial or detrimental to devices,[4] and will be discussed at length in Chapter 4. Also, this reduced symmetry causes difficulties in defining surface energies of the inequivalent N- and Ga-terminated *c* planes, which will be discussed in Chapter 3.

For reference, we list structural parameters and band gaps for GaN, AlN, InN, and BN in Table 1.1. The details of how the "HSE" columns are calculated are expanded upon in Section 1.2.4, and experimental data is provided where available.

### 1.1.2 Nitrides for light emitters

The most widespread application of III-nitride materials to date is for optoelectronics. GaN-based light-emitting diodes (LEDs) are a multi-billion-dollar industry based on applications including high-efficiency general lighting and as backlights for displays. In addition, III-nitride laser diodes (LDs) are used in high density optical storage, and open the door for exciting applications in lighting and projector technology.

As we see from Table 1.1, AlN, GaN and InN have direct band gaps ranging from 6.1 eV for AlN, to 3.5 eV for GaN, to 0.7 eV for InN. In principle, since they also have the same ground-state structure, these materials can be alloyed together to access any band gap in between; therefore, alloys could be made that absorb and emit light from 6.1 eV [203 nm wavelength, middle ultraviolet (UV)] to 0.7 eV (1770 nm wavelength, near infrared). Only using InN and GaN, InGaN alloys have band gaps which cover the visible part of the electromagnetic spectrum. However, in practice, the large lattice

**Table 1.1:** Parameters for the III-nitrides calculated with HSE (Section 1.2.4) compared with experimental data.

| | Property | HSE (this work) | Exp. |
|---|---|---|---|
| GaN | $a$ (Å) | 3.205 | 3.189[5] |
| | $c$ (Å) | 5.200 | 5.185[5] |
| | $u$ | 0.377 | 0.377[6] |
| | $E_g$ (eV) | 3.496 | 3.510[5] |
| AlN | $a$ (Å) | 3.099 | 3.112[5] |
| | $c$ (Å) | 4.959 | 4.982[5] |
| | $u$ | 0.382 | 0.382[6] |
| | $E_g$ (eV) | 6.044 | 6.250[5] |
| InN | $a$ (Å) | 3.587 | 3.545[5] |
| | $c$ (Å) | 5.762 | 5.703[5] |
| | $u$ | 0.380 | – |
| | $E_g$ (eV) | 0.646 | 0.780[5] |
| wz-BN | $a$ (Å) | 2.52 | 2.55[7–9] |
| | $c$ (Å) | 4.17 | 4.20-4.23[7–9] |
| | $u$ | 0.374 | – |
| | $E_g^{\text{indirect}}(\Gamma \rightarrow \text{K})$ (eV) | 6.46 | – |
| | $E_g^{\text{direct}}$(near-$\Gamma$) (eV) | 10.13 | – |
| h-BN | $a$ (Å) | 2.50 | 2.50–2.52[10] |
| | $c$ (Å) | 6.85 | 6.65–6.68[10] |
| | $E_g^{\text{indirect}}(\sim \text{H} \rightarrow \sim \text{M})$ (eV) | 5.65 | 3.8–5.9[11–15] |
| | $E_g^{\text{direct}}$(K) (eV) | 6.21 (M), 6.08 (K) | 5.2–5.97[11–15] |

mismatch between InN and GaN (about 12%) results in poor miscibility,[16] and limits the experimentally achievable alloys. Commercially, high-efficiency blue and violet (low In content) LEDs and LDs are now widespread, and are used to pump phosphors to produce white light. Currently, there is significant work to improve efficiency of devices in the green part of the spectrum (higher In content) in an attempt to remove the need for phosphors to make white light (direct mixing of light from red, green, and blue LEDs), as well as open the door for highly efficient red-green-blue LD displays.[17] Moving to AlGaN alloys opens up the possibility for light emitters in the deep UV, which have applications in even higher density optical storage, water purification, general disinfection, and spectroscopy.

In addition to simply accessing a wide range of wavelengths through alloying, the fact that the III-nitride materials share the wz structure allows the growth of thin-film heterostructures. The huge range in accessible band gaps (and band offsets) allows the engineering of heterostructures for a wide range of devices. Carriers can be confined in quantum wells with barriers for efficient light emission. Also light can also be confined with cladding layers of different refractive indexes, allowing laser structures to be grown.

Finally, a key property of GaN is that it can be doped *n*- and *p*-type, meaning that, depending on the impurity atoms added, it can conduct via electrons *or* holes. This means that *pn* junctions can be fabricated to inject both holes and electrons into the quantum wells, where they can recombine to emit light. The feasibility of "ambipolar

doping" is a rare quality for a wide-band-gap semiconductor, and necessary for high-efficiency light emitters.

### 1.1.3 Nitrides for transistors

In addition to optoelectronics, another promising application of III-nitride materials is for high frequency, high power transistors for devices such as microwave- and radio-frequency amplifiers, and power electronics for switching between alternating and direct current.[18] The large band gaps of GaN and AlGaN alloys result in significantly higher breakdown voltages than in traditional devices based on Si/Ge, InP, or GaAs.[18] As with optoelectronic devices, the significant band-gap difference between AlN and GaN (and their shared wurtzite ground state) means that heterostructures with AlGaN or AlInN barriers confining carriers to GaN channels can be grown.[18] Again, because of the possibility to dope GaN $n$- and $p$-type, these carriers can be electrons or holes, which allows the fabrication of bipolar transistors (involving $n$ and $p$ material),[19] increasing the range of possible applications. GaN also has relatively high conductivity compared to other wide-band-gap materials such as SiC.[18]

Another property of III-nitrides that is advantageous for the fabrication of electronic devices is the fact that the materials have large spontaneous and piezoelectric polarization moments due to the reduced symmetry of the wurtzite structure (discussed in detail

in Chapter 4). These result in electric fields in heterostructures which can be used to enhance carrier confinement.

### 1.1.4 Overcoming the limitations in the III-nitride system: Motivation for better understanding

Despite the plethora of attractive material parameters of the III-nitrides there are several issues that significantly limit the efficiency of devices and range of possible applications. Arguably the most detrimental issue is the lack of widely available, cost-effective substrates for the growth of films and devices.[20] Despite the significant advances in bulk GaN growth,[21] most III-nitride devices are grown heteroepitaxially on sapphire, SiC, or Si. Lattice and thermal-expansion-coefficient mismatch between the substrate and the III-nitride film, as well as other processes such as island coalescence,[22] result in residual stresses in the films. In addition, the significant lattice mismatch between GaN, InN, and AlN results in strain between layers of heterostructures. Because strain energy increases with thickness[23] and mismatch strain increases with alloy concentration, III-nitride device designs are limited in the thickness and alloy-content of barriers, cladding layers, and quantum wells. Therefore the range of possible wavelengths of optoelectronic devices is reduced significantly from the range expected by the band gaps. Also the ef-

ficiency of LDs is reduced due to less optical confinement, and the carrier confinement and two-dimensional electron gas (2DEG) density in transistors is bounded.

Stress will alter the electronic structure of the materials, so it is necessary to be able to quantify these effects for accurate device design. In Chapter 2 we explore the influence of strain on the effective mass of carriers in GaN and AlN, a parameter that is tied to the conductivity of the material. This work has been published in Ref. 24.

Residual stresses also degrade material quality. Dislocation concentrations of up to $10^8$–$10^{10}$ cm$^{-2}$ are common (compared to the near-perfect Si and GaAs films with less than $10^4$–$10^6$ cm$^{-2}$, see Ref. 25). In addition, films under tensile strain can crack if the strain energy is sufficient. We explore the propensity for AlN, GaN, and AlGaN to crack on different crystallographic planes in Chapter 3. This work is in preparation to be published in Ref. 26.

Significant work has been done to overcome the issue of residual strains in III-nitride films. As mentioned above, the growth of bulk crystals for substrates using techniques such as hydride vapor phase epitaxy (HVPE), ammonothermal growth,[21] and sodium flux growth[27] is showing rapid progress. Also, selective area growth (SAG) followed by lateral overgrowth by HVPE is being used to produce high-quality GaN for laser applications.[25] Similarly, several growth techniques including SAG have been used to create structures such as nanowires that avoid many of the thickness and alloy-content limitations of epitaxial layers and can be grown dislocation-free on nonnative substrates.[28] In

all of these cases, growth involves multiple growth planes, and the resulting morphology will be determined by the relative growth rates of the planes present. One of the fundamental parameters that determines the stability and growth properties on a given plane is the absolute surface energy of the plane. In Chapter 3 we report calculations of such surface energies for the most technologically important planes in GaN. This work has been published in Ref. 29.

As mentioned previously, the relatively low symmetry in the $c$ direction in the wurtzite structure results in polarization moments in that direction. The residual stresses discussed above will contribute to a piezoelectric polarization moment, but even in the absence of such strain, wurtzite crystals have spontaneous polarization moments in the $c$ direction.[4] Polarization differences between layers of heterostructures allow for strong carrier confinement and the formation of 2DEGs with high densities at AlGaN/GaN interfaces, exploited in high electron mobility transistors (HEMTs).[30] The effect of polarization can also be detrimental; for example causing the quantum-confined Stark effect in quantum wells of LEDs, which reduces radiative recombination rates and shifts the emission wavelength.[31] In both HEMT and LED devices, accurate values of the magnitudes of the spontaneous and piezoelectric polarization constants are required for a fundamental understanding and design of III-nitride devices, as discussed in Chapter 4. This work is in preparation for publications.[32]

## 1.2 Methods for determining properties of III-nitride materials from first principles

The properties of any solid-state system are determined, to a large extent, by the electronic structure of the valence electrons; these electrons participate in bonding and determine the electronic, elastic, and magnetic response of a material. The tools that we will use to explore the properties of III-nitride materials are first-principles electronic structure calculations based on density functional theory (DFT). The idea behind "first principles" (also referred to as *ab initio*) calculations is to start from the governing equation in (nonrelativistic) solid-state physics, the Schrödinger equation, and solve for the electronic structure and materials properties. In principle, this means that no empirical parameters are necessary, and the techniques can not only be used to explain and interpret experimental observations, but also to make predictions about systems and properties for which there is no experimental data available. In practice, however, a number of simplifications and assumptions must be made in the process of the calculations, some of which benefit from experimental input. These will be described in detail below. The discussion below is far from complete, however; the interested reader is referred to the reading list in Appendix A.

### 1.2.1 The many-body problem

One of the interesting aspects of solid-state physics is that the microscopic equation (many-body Schrödinger equation) that governs the macroscopic behavior of a system of electrons and nuclei, is known. Neglecting relativistic effects and the influence of magnetic fields, we can write it as (in Hartree atomic units)[33]

$$\begin{aligned}\Bigg[ -\frac{1}{2}\sum_i \nabla_i^2 + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|} - \sum_{i,I}\frac{Z_I}{|\mathbf{r}_i-\mathbf{R}_I|} - \sum_I \frac{1}{2M_I}\nabla_I^2 \\ + \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J}{|\mathbf{R}_I-\mathbf{R}_J|}\Bigg]\Psi(\{\mathbf{r}_i\},\{\mathbf{R}_i\};t) = i\frac{\partial\Psi(\{\mathbf{r}_i\},\{\mathbf{R}_i\};t)}{\partial t},\end{aligned} \tag{1.1}$$

where lower-case subscripts denote a given electron, upper-case subscripts denote a given nucleus, the $\mathbf{r}$'s are the electron positions, the $\mathbf{R}$'s are the nuclear positions, $M_I$ is the nuclear mass, and $Z$ is the nuclear charge. The first term on the left-hand side of Eq. (1.1) is the kinetic energy operator for the electrons, the second term is the coulomb interaction between electrons, the third term is the potential energy from the electron-nuclei interaction, the fourth term is the kinetic energy of the nuclei, and the final term represents the interactions between the nuclei. The properties of the system are determined by solving for the wave function $\Psi$. The solution to Eq. (1.1) will determine the material's atomic structure, whether it is metallic or insulating, its conductivity, whether or not it is superconducting, and so on.

Commonly, the so-called Born-Oppenheimer (BO) approximation is applied, where we take $M_I \to \infty$; this means that the kinetic energy of the nuclei goes to zero, so

the nuclei are treated classically.[33] Therefore, for the quantum-mechanical problem, the nuclei are treated as a fixed external potential and the electronic structure is determined for this fixed potential. Once the electronic structure is known, forces on the atoms can be determined using the Hellmann-Feynman theorem, and the atomic structure can be optimized. If the influence of the vibrational properties of the ions on the electronic structure is important, it is usually added using perturbation theory in the electron-phonon coupling.[33]

The total energy is the expectation value of the Hamiltonian

$$E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \langle \hat{T} \rangle_e + \langle \hat{V} \rangle_{ee} + \langle \hat{V} \rangle_{eI} \tag{1.2}$$

where we have written the electron kinetic energy operator [first term in Eq. (1.1)] as $\hat{T}$ and the potential energy operators as $\hat{V}$, where the subscript designates the interaction, $e$ for electron and $I$ for ions. With the BO approximation, the nuclei-nuclei and electron-nuclei interactions [third and fourth term in Eq. (1.2)] are given simply by classical electrostatics. The relative simplicity of Eqs. (1.1) and (1.2) conceals the fact that finding a solution is computationally intractable for more than a handful of electrons and nuclei, too few to describe most properties of bulk materials. Significant difficulty is due to the electron-electron interaction term, for which, as we see from Eq. (1.1), the number of of terms increases rapidly with number of electrons, significantly complicating the task of solving the Schrödinger equation.

### 1.2.2 Density functional theory

DFT is a method used to make this problem of calculating the electronic structure of a many-body systems computationally tractable. The basic idea behind DFT is to treat the electron *density*, instead of the explicit electron wave functions, as the defining variable of the system. Of course, the electron density, $n(\mathbf{r})$, is related to the many-body wave function through the relation

$$n(\mathbf{r}) = \frac{1}{N}\langle\Psi(\{\mathbf{r}_i\})|\sum_{i=1}^{N}\delta(\mathbf{r}-\mathbf{r}_i)|\Psi(\{\mathbf{r}_i\})\rangle. \tag{1.3}$$

where $N$ is the total number of electrons.

Hohenberg and Kohn[34] proved that all (ground- and excited-state) properties of a system in a given external potential are uniquely defined by the ground-state density of the system. This is not obvious since one might imagine that information about the system is lost when the integral $\langle\Psi|\Psi\rangle$ is taken. Also, Hohenberg and Kohn proved the additional theorem that for any external potential, there exists an energy functional of density, $E[n(\mathbf{r})]$, whose global minimum (over all $n$ that represent $N$ electrons) is the ground-state energy of the system, and the density that results in this functional being minimized is the exact ground-state density.[34]

The Hohenberg-Kohn theorems put DFT on a solid theoretical basis by showing rigorously that the (ground-state) electron density is sufficient to describe any property of a system. Even without any further considerations, the idea of dealing with the

density instead of the wave functions is an attractive one. Each many-body wave function depends on the positions of all the electrons in the system, meaning that for $N$ electrons, it has $3N$ variables. Imagine solving for such a wave function numerically on a grid of $M \times M \times M$ points and storing the solution. You would have to store $M^{9N}$ values of $\Psi(\{\mathbf{r}_i\})$. Therefore, even storing the solution (assuming you could solve the many-body Schrödinger equation) becomes prohibitively expensive for even a handful of electrons. For example, a wave function of 10 electrons on a 10×10×10 grid (assuming it takes one byte to store the value of $\Psi$ for a single grid point and one of the three position coordinates of a single electron) it would take $10^{78}$ terabytes to store the wave function. If we consider the density, however, $n(\mathbf{r})$ only has three spatial variables for *any* number of electrons; solving on the grid requires storing only $3M^3$ values, which scales much more favorably for a large number of electrons. For the example of 10 electrons on a 10×10×10 grid, storing the wave function would only require three kilobytes.

Though in principle the density treatment is attractive, the Hohenberg-Kohn theorems do not give any practical information about how to obtain the ground-state density of a system, or what $E[n(\mathbf{r})]$ might look like. It is far from obvious how to cast Eq. (1.2) in terms of densities. These practical details were explored in a subsequent publication by Kohn and Sham (KS).[35] In KS-DFT, the many-body-interacting system of electrons is replaced by an auxiliary system of noninteracting particles in an effective potential chosen such that it reproduces the ground-state density of the many-body system.

Starting from this *ansatz*, we attempt to write an energy functional of density for such a system in terms of *single* particle wave functions $\psi(\mathbf{r})$. Below we suppress explicit spin indices to avoid cluttering the notation. The derivation will follow that outlined in Ref. 33. The density for $N$ single-particle wave functions is given by

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\psi_i(\mathbf{r})|^2. \tag{1.4}$$

From simple quantum mechanics, the kinetic energy of $i$ single-particle wave functions is (in Hartree atomic units)

$$T_e^{\mathrm{s}} = \frac{1}{2} \sum_{i=1}^{N} \int d\mathbf{r} |\nabla \psi_i(\mathbf{r})|^2, \tag{1.5}$$

where we have designated it with "s" for "single particle" to differentiate it from the full many-body kinetic energy. Note that $T_e^{\mathrm{s}}$ is not explicitly a functional of the density, but of the single-particle wave functions. The kinetic energy is not a smoothly varying function of electron density, and therefore it turns out to be difficult to construct a kinetic energy functional of density (though, of course, it should be possible in principle). As is shown below, this does not turn out to be an issue.

Again, since we are operating under the BO approximation, we neglect the kinetic energy of the nuclei. Since we consider the potential of the nuclei as a fixed external potential, we write the nucleus-electron interaction from classical electrostatics as

$$E_{\mathrm{eI}} = \int d\mathbf{r} V_{\mathrm{ext}}(\mathbf{r}) n(\mathbf{r}), \tag{1.6}$$

and the nucleus-nucleus interaction is a constant with respect to the electron density [as in Eq. (1.1)]

$$E_{\mathrm{II}} = \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}. \tag{1.7}$$

We write one part of the electron-electron interaction as the Hartree energy of a classical charge density interacting with itself

$$E_{\mathrm{ee}}^{\mathrm{s}} = \frac{1}{2}\int d\mathbf{r}d\mathbf{r}'\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \tag{1.8}$$

The KS energy functional is

$$E_{\mathrm{KS}} = T_e^s + E_{\mathrm{ee}}^{\mathrm{s}} + E_{\mathrm{eI}} + E_{\mathrm{II}} + E_{\mathrm{XC}}. \tag{1.9}$$

Here we have introduced a new contribution to the energy, the exchange-correlation (XC) energy $E_{\mathrm{XC}}$. This term is defined as the contribution to the total energy from any many-body effects not included in the noninteracting kinetic energy and Hartree energy. The cynical reader will point out that we have, up to now, made no progress; we have simply placed all of the unknown many-body effects into a yet undefined term. However, the form of Eq. (1.9) is actually a very useful one, as we will demonstrate below.

The idea behind the KS formulation was to produce an auxiliary, effectively non-interacting system with the same ground-state density as the many-body treatment. According to the Hohenberg-Kohn theorems, the density that minimizes the exact energy density functional is the exact many-body ground-state density; therefore we minimize our energy functional in Eq. (1.9) with respect to the density. To do this, we

take the functional derivative of $E_{\rm KS}$ with respect to the complex conjugate of the $i$th single-particle wave function

$$\frac{\delta E_{\rm KS}}{\delta \psi_i^*(\mathbf{r})} = \frac{\delta T_{\rm e}^{\rm s}}{\delta \psi_i^*(\mathbf{r})} + \left( \frac{\delta E_{\rm ee}^{\rm s}}{\delta n(\mathbf{r})} + \frac{\delta E_{\rm eI}}{\delta n(\mathbf{r})} + \frac{\delta E_{\rm XC}}{\delta n(\mathbf{r})} \right) \frac{\delta n(\mathbf{r})}{\delta \psi_i^*(\mathbf{r})} = 0, \tag{1.10}$$

and enforce the orthonormalization constraint: $\langle \psi_i | \psi_j \rangle = \delta_{i,j}$. Note that in Eq. (1.10) the $E_{\rm II}$ term is absent since it is a constant, and its derivative with respect to the wave functions is zero. The functional derivatives of the energies in the parentheses can be converted into potentials. We can identify the functional derivative of the kinetic energy term as

$$\frac{\delta T_{\rm e}^{\rm s}}{\delta \psi_i^*(\mathbf{r})} = -\frac{1}{2}\nabla^2 \psi_i(\mathbf{r}). \tag{1.11}$$

Also, the derivative of the density with respect to the conjugate of the $i$th wave function is simply that wave function, $\psi_i(\mathbf{r})$. If we use the Lagrange multiplier method to implement the orthonormality condition, the minimization Eq. (1.10) can be written as

$$\left( -\frac{1}{2}\nabla^2 + V_{\rm ee}^{\rm sp} + V_{\rm eI} + V_{\rm XC} \right) \psi_i(\mathbf{r}) = \left( -\frac{1}{2}\nabla^2 + V_{\rm KS} \right) \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}), \tag{1.12}$$

where we have written the functional derivatives above as potentials, $V$. We see that for each single-particle wave function, we have a single-particle Schödinger equation (called a Kohn-Sham equation) with an effective KS potential chosen such that the sum of the solutions to Eq. (1.12) for all $i$ gives the exact, many-body ground-state density. Of course, the equations above demonstrate that $V_{\rm KS}$ also depends on the density; therefore a solution to Eq. (1.12) must be solved selfconsistently.

### 1.2.3 Traditional forms for the exchange-correlation potential and the band-gap problem

In principle, Eq. (1.12) will give the exact ground-state density. However, nowhere in the analysis did we define the so called XC potential, $V_{\mathrm{XC}}$. In fact, there is no known form for the exact $V_{\mathrm{XC}}$. Kohn and Sham did provide an approximate version for $V_{\mathrm{XC}}$ called the local density approximation (LDA).[35] $V_{\mathrm{XC}}^{\mathrm{LDA}}$ approximates the XC at a given point $\mathbf{r}$ as that of the uniform electron gas with the same density as the KS density $n(\mathbf{r})$ at $\mathbf{r}$. The XC potential for the uniform electron gas can be calculated exactly using quantum Monte Carlo techniques, and tabulated with respect to density. Therefore, it is a simple, local (only depending on the density at a given point), smooth function of the density. LDA has been amazingly successful at capturing many properties of many materials, and is attractively intuitive and simply parametrized.[36,37] Further developments, called the generalized gradient approximation (GGA), included terms dependent on the gradient of the density to try to better capture the difference between the homogeneous electron gas and the inhomogeneities of a real material.[38]

One of the most significant problems with LDA and GGA is that they significantly underestimate the band-gap of insulators (given by the energy difference between the highest occupied single-particle state and the lowest unoccupied), often by 50 % or more. This so-called "band-gap problem" brings up the subtle issue of the meaning of

the eigenvalues and eigenfunctions of the KS equation. All that was guaranteed by the derivation above was that the KS ground-state density and total energy correspond to the exact many-body ground-state density and total energy (for an exact $V_{\mathrm{XC}}$); there was no mention of the eigenvalues or even the single-particle wave functions having any physical significance. It can be shown that the highest occupied eigenvalue correctly gives the negative ionization energy,[39] but the others do not represent the true energies for adding or removing electrons from the system. However, it has been shown empirically, and by adding quasiparticle interaction perturbatively,[40] that the single-particle eigenvalues are a good approximation to the true quasiparticle energies of the electrons in a solid.

Even if we trust that the single-particle wave functions and eigenvalues have some physical significance, it is still unclear that the band gap is physical since it is an excited state property. Of course, the Hohenberg-Kohn theory states that any excited state properties can be determined from the ground-state density, but again, in the KS analysis, all that we are guaranteed is the correct ground-state density and there is no obvious practical way to take advantage of Hohenberg-Kohn and obtain the excited state properties from this density. We make the assumption that for a macroscopic material (which is essentially infinite on the scale of the atoms and electrons), the neutral excitation of a single (or very few) electrons is an infinitesimal perturbation to the system. Therefore, we consider the electronic structure of such an excited system as basically

that of the ground state, with the excited electron occupying the lowest unoccupied conduction-band wave function.

### 1.2.4 Hybrid functionals

We now return to the band-gap problem. If we assume that a better XC approximation will give us a closer result to the true quasiparticle band structure, the task is to improve $V_{\mathrm{XC}}$. An attempt at a more accurate functional is the concept of hybrid functionals, initially developed by Becke.[41] The idea initiates from a form of the XC energy that results from adiabatic connection theory,[42] where a system of noninteracting particles (charge neutral) is slowly transitioned to a fully interacting system by adiabatically increasing the charge:

$$E_{\mathrm{XC}} = \int_0^1 V_{\mathrm{XC}}(\gamma) d\gamma, \tag{1.13}$$

where $\gamma$ is the charge of the particles and $V_{\mathrm{XC}}$ is the exact XC potential. The lower limit of the integral, $V_{XC}(\lambda = 0)$, is the interaction of a system of uncharged particles, and corresponds to the Hartree-Fock "exact" exchange energy of the system.[41] This energy is given by

$$E_{\mathrm{X}}^{\mathrm{HF}} = -\frac{1}{2} \sum_{i,j,\sigma} \int d\mathbf{r} d\mathbf{r}' \psi_{i,\sigma}^*(\mathbf{r}) \psi_{j,\sigma}^*(\mathbf{r}') \frac{1}{|\mathbf{r} - \mathbf{r}'|} \psi_{j,\sigma}(\mathbf{r}) \psi_{i,\sigma}(\mathbf{r}') \tag{1.14}$$

where the sum goes over a single spin $\sigma$. This term looks very similar to the Hartree energy described above for KS theory (for a single spin), though because the integral

over $\mathbf{r}'$ involves both $\psi_i$ and $\psi_j$ (the terms are "exchanged" compared to the Hartree energy), it cannot be expressed in terms of densities.

The $\gamma = 1$ limit corresponds to the fully interacting system, which we have approximated previously with LDA or GGA. Therefore, even though we do not know the exact form of $V_{\mathrm{XC}}$, we can see (from the mean value theorem) that a better approximation to the integral than LDA/GGA alone is some linear combination of Hartree-Fock exchange and LDA exchange and correlation.[41] The XC energy in these hybrid functionals is given by

$$E_{\mathrm{XC}} = \alpha E_{\mathrm{X}}^{\mathrm{HF}} + (1-\alpha) E_{\mathrm{X}}^{\mathrm{LDA/GGA}} + E_{\mathrm{C}}^{\mathrm{LDA/GGA}}, \tag{1.15}$$

where $\alpha$ is an adjustable "mixing" parameter.

In this work we use the hybrid functional developed by Heyd, Scuseria, and Ernzerhof[43] (HSE), which has a second parameter, $\sigma$ that screens the long-range part of the Hartree-Fock exchange

$$E_{\mathrm{XC}} = \alpha E_{\mathrm{X}}^{\mathrm{HF,SR}}(\sigma) + (1-\alpha) E_{\mathrm{X}}^{\mathrm{GGA,SR}}(\sigma) + E_{\mathrm{X}}^{\mathrm{GGA,LR}}(\sigma) + E_{\mathrm{C}}^{\mathrm{GGA}}, \tag{1.16}$$

where the Hartree-Fock exchange is now replacing only the short-range exchange of the GGA XC. [Specifically, the GGA functional developed by Perdew, Burke, and Ernzerhof[38] (PBE) is used for the GGA XC part of HSE.]

In general, hybrids provide an improved description of the electronic structure of semiconductors and isolators, including parameters such as effective masses and the po-

sition of semicore states in the band structure, compared to LDA/GGA.[44] Arguably the main benefit is that they produce much more accurate gaps for insulators and semiconductors.[45] The gap in a given material has an approximately linear dependence on the $\alpha$ parameter in Eq. (1.16). The default value for HSE is 0.25,[46] but there is no good reason why it should be constant for all materials. Often, in the course of the work reported here, we use different values of $\alpha$ to obtain experimentally accurate band gaps.

Beyond the empirical success of hybrids, the specific reasons for the improvement in the band gaps is subtle. Without getting into too many details, we consider a few properties of such functionals that assist intuitive understanding. From a completely naive perspective, Hartree-Fock theory tends to overestimate band gaps and LDA/GGA underestimates band gaps, so one would expect combining the two meets in the middle. This, however, does not account for the accuracy of band gaps over a large range of materials, even without tuning the mixing parameter. A more careful analysis revels two qualitative aspects of hybrids that play a role in their correction to the electronic structure.

First, as mentioned above, the Hartree-Fock exchange functional involves the wave functions explicitly as opposed to the density. Also mentioned above, the XC functional was expected to be a discontinuous function because part of it was derived from the missing many-body contribution to the kinetic energy. LDA and GGA are smoothly varying functions with density, and one of the main reasons for the band-gap problem is

the lack of a "derivative discontinuity," where the XC energy should increase discontinuously when a band (such as the highest lying valence band) is filled and a separate band (such as the lowest lying conduction band) begins to be occupied. Such discontinuity suggests that a smoothly varying function of the density is not sufficient to capture the energy of gaps in the electronic structure; it would be expected that explicit use of the wave functions would better capture the discontinuous behavior as different bands become occupied.

The second issue with LDA and GGA that hybrids influence is the self-interaction error. Consider the Hartree energy in Eq. (1.8); we can write it out explicitly in terms of single-particle wave functions (now including spin)

$$E_{\mathrm{ee}}^{\mathrm{s}} = \frac{1}{2} \sum_{i,j,\sigma_i,\sigma_j} \int d\mathbf{r} d\mathbf{r}' \psi_{i,\sigma_i}^*(\mathbf{r}) \psi_{j,\sigma_j}^*(\mathbf{r}') \frac{1}{|\mathbf{r} - \mathbf{r}'|} \psi_{i,\sigma_i}(\mathbf{r}) \psi_{j,\sigma_j}(\mathbf{r}'). \tag{1.17}$$

We see that the term in the sum where $i = j$ is the interaction between the charge density of a state with itself. In Hartree-Fock, this term is exactly cancelled for each $i$ by the exchange term [Eq. (1.14)]. This spurious self interaction causes LDA and GGA to over-delocalize the charge density, or spuriously favor charge densities that are more delocalized. Adding in some amount of Hartree-Fock exchange partially cancels the self interaction. This partial cancellation of the self interaction also allows hybrid functionals to better describe the energetics of localized systems where self interaction is significant.

The above discussion is only an intuitive guide. In practical terms, hybrid functionals have been shown to well describe the electronic and atomic structure of III-nitride materials.[44,47,48] The increased accuracy does come at a computational cost as the calculation of the Hartree-Fock exchange results in an increase in computation time for a given calculation by about an order of magnitude. Even so, the HSE functional can be applied to systems of order 100 atoms in the nitrides, which is often sufficient to capture macroscopic properties when we take into account the periodic nature of crystalline systems.

### 1.2.5 Practical aspects of density functional theory calculations

The above discussion dealt with general theoretical aspects of DFT and hybrid DFT; here I will make a few comments on important aspects of running practical calculations that will be referenced throughout the document. All DFT calculations were carried in the Vienna *ab initio* simulations package (VASP).[49] The XC functionals used were PBE-GGA[38] and HSE,[46] as discussed above.

The first practical aspect of solving the KS equation numerically is that we must choose a basis set in which to parametrize the wave functions. The basis set is chosen for numerical convenience depending on the application; for our calculations, we expand

the wave functions in plane waves. This choice has several advantages for the types of calculations that we perform. Checking convergence of the basis set is simple since the accuracy of the expansion increases monotonically with the highest-energy plane wave included, specified by an energy cutoff parameter. This is not the case with all basis sets; increasing the size of a gaussian basis set, for example, does not guarantee systematic convergence,[33] complicating convergence checks. Also, integrals with plane waves are relatively simple to perform numerically using fast Fourier transforms. One issue with a plane wave basis set, however, is that the number of plane waves necessary to describe a rapidly varying potential, such as that of atomic nuclei and core electrons, is prohibitively large. This is overcome by replacing the wave functions of the core electrons and nuclei with a more smoothly varying "pseudopotential" that reproduces all of the long-range behavior of the full, all-electron potential. In our calculations we use the Projector Augmented Wave (PAW)[50] method to generate such pseudopotentials.

Because of the use of plane waves, all of our calculations are performed under periodic boundary conditions in all directions. This is beneficial for bulk calculations, as it allows the simulation of a macroscopic solid (practically infinite on the scale of atoms) with a relatively small unit cell. Systems with lower dimensional periodicity, such as surfaces, nanowires, isolated point defects, or individual molecules can also be calculated under periodic boundary conditions, where calculations are performed on"supercells" that are larger than the primitive unit cell of the material, and can include vacuum. If enough

vacuum and/or bulk material is included, supercells will spatially isolate structures (such as surfaces or point defects) from their periodic images, avoiding spurious interactions.

With plane waves and periodic boundary conditions, the KS equation can be solved efficiently in reciprocal space. Many quantities of importance, such as the total energy, therefore, can be expressed as a sum over **k** vectors in the first Brillouin Zone (BZ) of the unit cell. The computation time scales with the number of **k** vectors at which the wave functions need to be calculated; therefore it is important to find the lowest number of $k$-points to calculate that will give the required accuracy. A naive idea would be to calculate over a regular grid in $k$-space, but we can do a lot better than that. It was shown, first by Baldereschi[51] and later by Chadi and Cohen[52] and Monkhorst and Pack (MP)[53] that for a given number of $k$-points, we can take advantage of the symmetry of the system to determine which $k$-points will give us the best approximation to a full integration over the BZ. Therefore, we simply have to define a mesh of points in three dimensions, and using the MP procedure, choose the **k** vectors to give the highest accuracy BZ integration for a given mesh density.

## 1.3 Conclusions

The III-nitrides are a technologically exciting class of materials for applications in light emitters and electronics. Motivated by the necessity for a better understanding

of the fundamental material properties, we have performed first-principles calculations based on DFT on bulk and supercells of AlN, GaN, InN, and BN. DFT is a powerful tool to calculate the electronic structure of solid-state systems in a computationally tractable way. In order to obtain accurate parameters such as band gaps we have used the HSE hybrid functional in this study, which provides a better approximation to the exact XC energy compared to traditional functionals such as LDA and GGA. In the following chapters, we will outline the methods, results, and interpretations of these calculations.

# Chapter 2

# The effect of strain on the effective mass of GaN and AlN

As mentioned in Chapter 1, III-nitride films are often subject to residual stresses from lattice mismatch or other sources. The electronic and transport properties of semiconductors can be significantly altered by applying stress. Carrier-mobility enhancement with tensile stress is well characterized in Si, and this effect is exploited in the design of Si-based transistors.[54]

Recent experiments on high electron mobility transistors (HEMTs) based on AlGaN/GaN heterostructures indicate an increase of 25% in sheet electron density, and an enhancement as large as 20% in the electron mobility with 0.16% tensile strain under applied biaxial stress[55] (see Section 2.1). The authors speculated that the increase

in mobility resulted from a decrease in the electron effective mass due to strain. This hypothesis motivated us to examine the variation of electron effective mass with strain in GaN and AlN by performing state-of-the-art first-principles calculations. We also provide an interpretation of our results in the context of $\mathbf{k} \cdot \mathbf{p}$ theory (explained in Section 2.2) and show that the variation of the interband momentum matrix elements with strain needs to be taken into account to produce accurate trends. Our approach illustrates the power of combining first-principles DFT calculations with semi-empirical $\mathbf{k} \cdot \mathbf{p}$ modeling.

In this Chapter we will make a few brief comments on stress and strain (Section 2.1), give an overview of $\mathbf{k} \cdot \mathbf{p}$ theory (Section 2.2), and then present our results for the effect of strain on the electronic structure of GaN and AlN [Section 2.3, reprinted excerpt with permission from Ref. 24, copyright (2013) AIP Publishing].

## 2.1 Stress and strain in epitaxial III-nitrides

In general, for small deformations from equilibrium of a crystal lattice, the stress ($\sigma$) and strain ($\epsilon$) are related by the symmetric fourth-rank elasticity tensor, $c_{ijkl}$:[4]

$$\sigma_{ij} = c_{ijkl}\epsilon_{kl}. \tag{2.1}$$

To make the notation more compact, we use Voigt notation to express this as a second-order tensor:[4]

$$\begin{array}{cccccccc} ij = & 11 & 22 & 33 & 23,32 & 13,31 & 12,21 \\ \Downarrow & \Downarrow & \Downarrow & \Downarrow & \Downarrow & \Downarrow & \Downarrow \\ \alpha = & 1 & 2 & 3 & 4 & 5 & 6 \end{array} \tag{2.2}$$

For hexagonal lattices (such as wurtzite), the tensor has five independent components:[4]

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ \cdot & c_{11} & c_{13} & 0 & 0 & 0 \\ \cdot & \cdot & c_{33} & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & c_{44} & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & c_{44} & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \frac{1}{2}(c_{11} - c_{12}) \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}. \tag{2.3}$$

In the linear elastic regime, stress and strain can be used interchangeably, with the understanding that they are directly proportional. However, care must be taken differentiating between stress and strain *states*. For example, the hydrostatic strain is defined as $\epsilon_1 = \epsilon_2 = \epsilon_3$, which is not necessarily the same as hydrostatic stress: $\sigma_1 = \sigma_2 = \sigma_3$ if the material is not isotropic (meaning $c_{11}$, $c_{22}$, and $c_{33}$ are not all equal). Also, there is *always* a difference between biaxial stress ($\sigma_1 = \sigma_2$, $\sigma_3 = 0$) and biaxial strain ($\epsilon_1 = \epsilon_2$, $\epsilon_3 = 0$), even for isotropic materials. Biaxial stress is a commonly occurring stress state in psuedomorphically grown epitaxial films, whereas biaxial strain

is difficult to achieve experimentally (though can be done in first-principles calculations and may be useful for calculating deformation potentials[56]).

In the discussion below we use the intuitive notation that $\alpha = 1 \rightarrow x$, $\alpha = 2 \rightarrow y$, and $\alpha = 3 \rightarrow z$.

## 2.2 $\mathbf{k} \cdot \mathbf{p}$ for exploring strain effects on electronic structure

An important tool for modeling the strain effects on band structures is the so-called $\mathbf{k} \cdot \mathbf{p}$ method. In the context of III-V semiconductors (wurtzite and zincblende), $\mathbf{k} \cdot \mathbf{p}$ usually refers to the specific semi-empirical parametrization that is outlined at the end of this section. However, the method is very general and based simply on perturbation theory applied to periodic lattices.

For periodic crystals, the external potential and all terms in the Hamiltonian have to be commensurate with the translational symmetry of the crystal. We know from Bloch's theorem[57] that the eigenfunctions for a translationally invariant Hamiltonian can be written in the form

$$\psi(\mathbf{k}, \mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) u(\mathbf{k}, \mathbf{r}), \tag{2.4}$$

where $\mathbf{k}$ is a wave vector of the crystal momentum, and the function $u(\mathbf{k}, \mathbf{r})$ is periodic with the lattice periodicity. The wave vector $\mathbf{k}$ is related to the eigenvalue of the translation operator in the $j$ direction, $T_j$, given by $\exp(i\mathbf{k} \cdot \mathbf{a}_j) = T_j$, where $\mathbf{a}_j$ is the corresponding lattice vector. If we substitute this form of the wave function into the single-particle Schrödinger equation with Hamiltonian $\hat{H}$ and multiply the result by $\exp(-i\mathbf{k} \cdot \mathbf{r})$, we get the Schrödinger-Bloch equation:[57]

$$\left(\hat{H} + \frac{\hbar}{m_e}\mathbf{k} \cdot \mathbf{p} + \frac{\hbar^2 k^2}{2m_e}\right) u(\mathbf{k}, \mathbf{r}) = E_{\mathbf{k}} u(\mathbf{k}, \mathbf{r}), \tag{2.5}$$

For each $\mathbf{k}$ vector, Eq. (2.5) has an infinite number of discrete solutions, forming a complete set; therefore, Bloch lattice functions for any given value of $\mathbf{k}$ can be expanded in terms of those of any other $\mathbf{k}$. Specifically, we can use perturbation theory to obtain the Bloch waves in the vicinity of $\mathbf{k} = 0$ by expanding in terms of functions at $\mathbf{k} = 0$.[57] This can be a very useful procedure, as much of the physics of direct-band-gap semiconductors, such as GaN, happen in the vicinity of the $\Gamma$ ($\mathbf{k} = 0$) point.

Taking the perturbing potential as

$$W_{mn} = \frac{\hbar}{m_e}\mathbf{k} \cdot \langle u_m(\mathbf{k} = 0; \mathbf{r})|\mathbf{p}|u_n(\mathbf{k} = 0; \mathbf{r})\rangle = \frac{\hbar}{m_e}\mathbf{k} \cdot \mathbf{p}_{mn} \tag{2.6}$$

we can write the second-order correction to the energy using nondegenerate Rayleigh-Schrödinger perturbation theory[57]

$$\Delta E_n = \frac{\hbar^2}{m_e^2} \sum_{m \neq n} \frac{|\mathbf{k} \cdot \mathbf{p}_{mn}|^2}{E_n(\mathbf{k} = 0) - E_m(\mathbf{k} = 0)}. \tag{2.7}$$

Therefore, to determine this correction, we need to calculate momentum matrix elements $\mathbf{p}_{mn}$ and energy gaps $E_n(\mathbf{k}=0) - E_m(\mathbf{k}=0)$ at the $\Gamma$ point. If we assume cubic symmetry ($|\mathbf{k}\cdot\mathbf{p}_{mn}|^2 = k^2|p_{mn}|^2$) and add the $k^2$ term from the right hand side of Eq. (2.5) we get the dispersion of band $n$ near $\Gamma$ as

$$E_n(\mathbf{k}) - E_n(0) = \frac{\hbar^2 k^2}{2m_e}\left(1 + \frac{2}{m_e}\sum_{m\neq n}\frac{|p_{mn}|^2}{E_n(0) - E_m(0)}\right). \tag{2.8}$$

Taylor expanding the energy around $\Gamma$ leads to an expression for the effective mass of band $n$, $m_n^*$, as[57]

$$\frac{m_e}{m_n^*} = 1 + \frac{2}{m_e}\sum_{m\neq n}\frac{|p_{mn}|^2}{E_n(0) - E_m(0)}. \tag{2.9}$$

We see from this analysis that the effective mass is proportional to the band gap, and inversely proportional to the momentum matrix element. We will discuss this further in Section 2.3.2.

A very useful version of the $\mathbf{k}\cdot\mathbf{p}$ formalism was developed by Kane[58] for the case of zincblende III-V semiconductors. The so-called Kane model uses a degenerate perturbation theory expansion with four basis functions corresponding to the three highest valence bands, which are derived from the anion $p$ states, and the lowest conduction band, which is derived from the cation $s$ states. The generalization for wurtzite III-nitrides is trivial[59,60] since the same four bands make up the valence-band maximum and conduction-band minimum at the $\Gamma$ point. We express the basis functions based on the states they are derived from ($S$ for cation $s$ states and $X$,$Y$, and $Z$ for the anion $p_x$,

$p_y$, and $p_z$) as

$$|iS\rangle,\ \frac{1}{\sqrt{2}}|X+iY\rangle,\ \frac{1}{\sqrt{2}}|X-iY\rangle,\ |Z\rangle. \tag{2.10}$$

We can therefore construct a $4\times 4$ Hamiltonian, neglecting spin and spin-orbit coupling

$$H_{4\times 4}=\frac{\hbar^2k^2}{2m_0}+\begin{pmatrix} E_c & -\frac{P_2}{\sqrt{2}}(k_x+ik_y) & \frac{P_2}{\sqrt{2}}(k_x-ik_y) & P_1k_z \\ -\frac{P_2}{\sqrt{2}}(k_x-ik_y) & E_v+\Delta_{\mathrm{cf}} & 0 & 0 \\ \frac{P_2}{\sqrt{2}}(k_x+ik_y) & 0 & E_v+\Delta_{\mathrm{cr}} & 0 \\ P_1k_z & 0 & 0 & E_v \end{pmatrix}, \tag{2.11}$$

where $E_c$ and $E_v$ are the conduction- and valence-band energies, $\Delta_{\mathrm{cf}}$ is the crystal field splitting, and $P_1$ and $P_2$ are momentum matrix elements:

$$P_1=\frac{\hbar^2}{m_e}\langle S|\frac{\partial}{\partial z}|Z\rangle,$$
$$P_2=\frac{\hbar^2}{m_e}\langle S|\frac{\partial}{\partial x}|X\rangle=\frac{\hbar^2}{m_e}\langle S|\frac{\partial}{\partial y}|Y\rangle.$$

We will calculate $\Delta_{\mathrm{cf}}$, $P_1$, and $P_2$ at $\Gamma$ from first principles (though experimental data is also often used) and solve the Hamiltonian in Eq. (2.11) to give the dispersion of the bands around $\Gamma$.

As mentioned above, one of the useful applications of $\mathbf{k}\cdot\mathbf{p}$ is to parametrize the effect of some strain state, $\epsilon_{ij}$, on the band structure. This has been derived by Bir and Pikus.[60] In their parametrization, the Hamiltonian matrix is expressed in terms of band parameters $A_i$ (see Ref. 60 for definitions) and deformation potentials $D_i$, $a_{cz}$, and $a_{ct}$.[60] Neglecting spin and spin-orbit coupling, the Hamiltonian and strain-dependent elements

are

$$H_{4\times4} = \begin{pmatrix} E_c & 0 & 0 & 0 \\ 0 & F & -K^* & -H^* \\ 0 & -K & G & H \\ 0 & -H & H^* & \lambda \end{pmatrix} \tag{2.12}$$

$$F = G = \Delta_{\text{cf}} + \lambda + \theta$$

$$\lambda = \frac{\hbar^2}{2m_e}[A_1 k_z^2 + A_2(k_x^2 + k_y^2)] + D_1\epsilon_{zz} + D_2(\epsilon_{xx} + \epsilon_{yy})$$

$$\theta = \frac{\hbar^2}{2m_e}[A_3 k_z^2 + A_4(k_x^2 + k_y^2)] + D_3\epsilon_{zz} + D_4(\epsilon_{xx} + \epsilon_{yy})$$

$$K = \frac{\hbar^2}{2m_e}A_5(k_x + ik_y)^2 + D_5(\epsilon_{xx} + 2i\epsilon_{xy} - \epsilon_{yy})$$

$$H = \frac{\hbar^2}{2m_e}A_6(k_x + ik_y)k_z + D_6(\epsilon_{zx} + i\epsilon_{yz})$$

$$E_c = E_v^{k=0} + \Delta_{\text{cf}} + E_g + a_{cz}\epsilon_{zz} + a_{ct}(\epsilon_{xx} + \epsilon_{yy}),$$

where $E_g$ is the zero-strain band gap and $E_v^{k=0}$ is the valence-band maximum at $\Gamma$, also at zero strain. The above parametrization may seem complicated, but once the band parameters (which consist of gaps and momentum matrix elements, similar to the Kane model) and deformation potentials are calculated, the effect of strain on gaps and dispersions can be included in semi-emperical calculations, such as Schrödinger-Poisson device simulations, in a very straightforward and computationally efficient implementation.

## 2.3 First-principles calculations of strained GaN and AlN

Returning to the case of GaN and AlN, we have used first-principles calculations to directly determine the effect of strain on the band dispersions, and determined the effect of our results on the $\mathbf{k} \cdot \mathbf{p}$ parametrization described above.

The effects of stress on the band gap and valence-band structure of III-nitrides have been thoroughly investigated, but there have been few studies on the effects of stress on electron effective mass. Gorczyca *et al.*[61] used density functional theory (DFT) to study the variation of effective mass with hydrostatic pressure in InN. Svane *et al.*[62] used quasiparticle calculations for GaN, but again focused only on hydrostatic pressure.

The DFT-HSE[43,46] (Section 1.2.4) calculations are performed using the VASP code,[49] implemented with projector augmented wave (PAW) potentials.[50] The Ga $d$ electrons were frozen in the core of the pseudopotential. Tests including the Ga $d$ electrons in the valence indicate that our conclusions remain unaffected. The hybrid functional approach produces accurate band structures and correctly captures the effects of strain on the band structure of nitrides.[56] We have found that it also produces equilibrium effective masses in agreement with experiment (although the reported values vary over a wide range)[5] and with quasiparticle $G_0W_0$ calculations[63] (see Table 2.1). We use an energy cutoff of 600 eV for the plane-wave basis set, a mixing parameter of 31 %, and a mesh

**Table 2.1:** Equilibrium (zero-strain) electron effective mass parallel and perpendicular to the $c$ direction in units of free electron mass compared to selected literature values.

| | Direction | Experiment[5] | $G_0W_0$[63] | HSE (this work) |
|---|---|---|---|---|
| GaN | $m_e^{\parallel}$ | $0.18-0.29$ | 0.19 | 0.19 |
| | $m_e^{\perp}$ | $0.18-0.29$ | 0.21 | 0.22 |
| AlN | $m_e^{\parallel}$ | $0.29-0.45$ | 0.32 | 0.30 |
| | $m_e^{\perp}$ | $0.29-0.45$ | 0.33 | 0.32 |

of 6×6×4 special **k**-points for the integrations over the Brillouin zone of the wurtzite primitive cell. The calculated band gaps and lattice constants are in good agreement with experimental values (see Table 1.1).

We have studied the variation of the band gap and electron effective mass in GaN and AlN under both hydrostatic and in-plane (i.e., perpendicular to the $c$ direction) biaxial stress. We approximate hydrostatic stress by imposing hydrostatic *strain*, i.e., changing the lattice parameters of the wurtzite cell equally ($\epsilon_{xx}$=$\epsilon_{yy}$=$\epsilon_{zz}$) and allowing the internal structural $u$ parameter to fully relax. The $c/a$ ratio of the wurtzite structure slightly decreases with increasing hydrostatic stress,[64] but the effect on the band gap is negligibly small for the relatively small strains considered here.[61] For the case of biaxial stress, the in-plane lattice parameters were changed equally, and the crystal was allowed to relax in the $c$ direction ($\epsilon_{xx}$= $\epsilon_{yy}$≠$\epsilon_{zz}$). In this configuration, the macroscopic theory of elasticity predicts that the strains should be related by $\epsilon_{zz}$=$-2(C_{13}/C_{33})\epsilon_{xx}$. Our

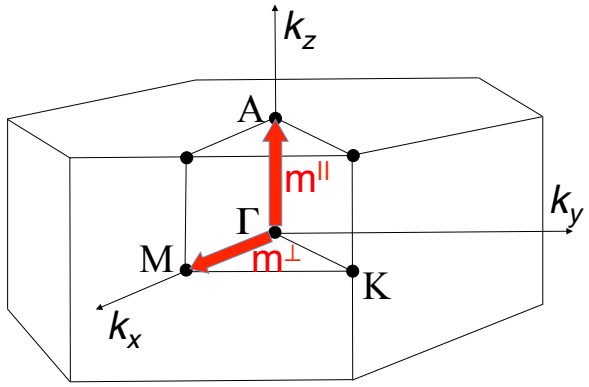


**Figure 2.1:** Brillouin zone of the wurtzite structure indicating high-symmetry directions. Arrows denote directions along which effective masses were calculated.

calculated values of $-2C_{13}/C_{33}$ (–0.564 for GaN and –0.571 for AlN) are within the range of reported experimental values[5] (–0.369 to –0.684 for GaN and –0.509 to –0.606 for AlN).

We determined the electron effective mass by analyzing the calculated band structure near Γ along the A direction in reciprocal space ($m_e^{\parallel}$, parallel to the $c$ direction of the crystal) and along the M and K directions ($m_e^{\perp}$, perpendicular to the $c$ direction), as indicated in Fig. 2.1. We found the electron effective mass in the M and K directions to be equal at zero strain, and to remain equal under the stress states explored here.

### 2.3.1 Effect of strain on the band gap

Figure 2.2 displays our results for the variation of band gap under imposed strain. Hydrostatic stress causes the band gap to increase linearly under compression, with a deformation potential $a_g = dE_g/d\ln V$ of −8.31 eV for GaN and −9.88 eV for AlN. These

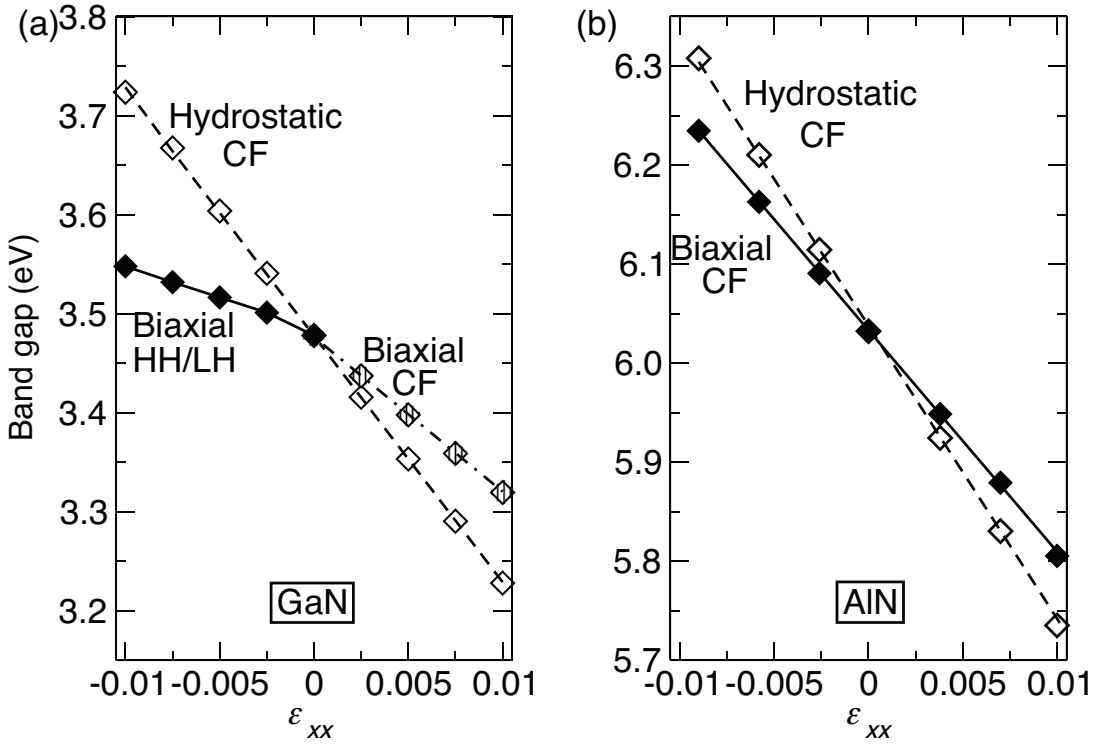


**Figure 2.2:** Dependence of band gap on in-plane strain for (a) GaN and (b) AlN under hydrostatic and biaxial stresses. The topmost valence band is specified in the label.

values fall within the range of reported experimental hydrostatic deformation potentials: –7.8 to –11.8 eV for GaN and –7.1 to –11.8 eV for AlN.[5]

For biaxial stress, the band gap also increases with compression, but the ordering of the topmost valence-band states is now affected. The relevant valence bands near Γ, denoted by their symmetry character,[65] are the heavy hole HH($\Gamma_9$) and light hole LH($\Gamma_7$) bands, and the crystal-field split-off band CF($\Gamma_7$). The HH and LH bands are degenerate because spin-orbit interaction (which is very small in the nitrides, 17.3 meV for GaN and 21.7 meV for AlN, Ref. 44) is not included here.

For GaN, the HH($\Gamma_9$) and LH($\Gamma_7$) bands are highest in energy under compressive strain. Under tensile biaxial stress the CF($\Gamma_7$) band becomes the highest-lying valence band. This leads to a kink between the linear regions in the band-gap-versus-strain plot [Fig. 2.2(a)], consistent with experimental observations.[66] The variation of band gap with volume ($dE_g/d\ln V$) for biaxial stress is $-10.90$ eV for the energy gap between the CF($\Gamma_7$) valence band and the conduction band, and $-4.29$ eV for the energy gap between the HH($\Gamma_9$)/LH($\Gamma_7$) valence band and the conduction band. In unstrained AlN, the highest energy band is the CF($\Gamma_7$) band, and it remains the topmost valence band over the entire strain range explored here [Fig. 2.2(b)], consistent with experimental observations.[67] For AlN under biaxial stress, $dE_g/d\ln V = -15.76$ eV. Values of $dE_g/d\ln V$ calculated here are within 8 % of the values derived from calculations in Ref. 56.

### 2.3.2 Effect of strain on the effective mass

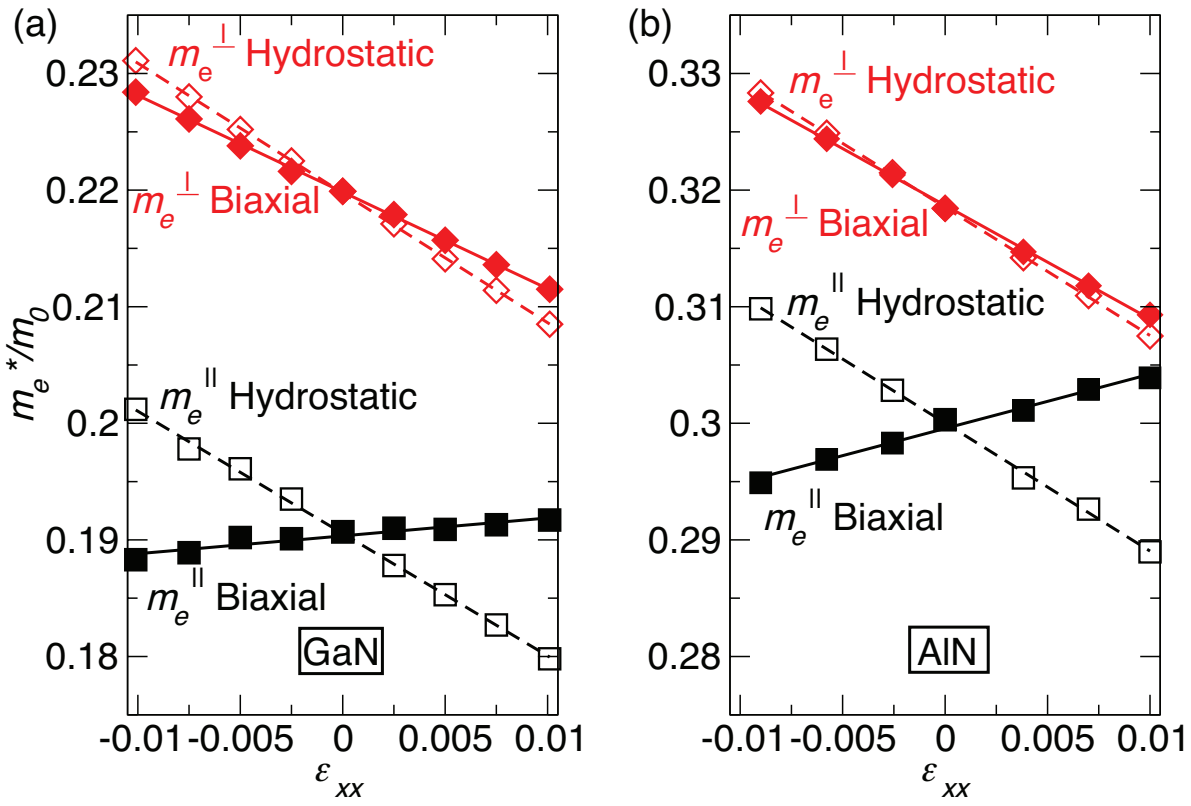


**Figure 2.3:** Dependence of effective mass on in-plane strain for (a) GaN and (b) AlN, for hydrostatic as well as biaxial stress. The parallel ($m^{\parallel}$) and in-plane ($m^{\perp}$) directions are indicated in Fig. 2.1.

Figure 2.3 shows the variation of the electron effective mass with strain in GaN and AlN. For the case of hydrostatic stress, tensile strain causes the effective mass parallel and perpendicular to the $c$ direction to decrease linearly in both GaN and AlN. The calculated variation of the effective mass with volume (i.e., under hydrostatic stress) for GaN, expressed as $d\ln m^*/d\ln V$, is $-1.7$ for $m_e^{\perp}$ and $-1.8$ for $m_e^{\parallel}$. These values are

in good agreement with those calculated by Svane *et al.*[62] using the quasiparticle, self-consistent *GW* approach: $-1.8$ for $m_e^{\perp}$ and $-1.9$ for $m_e^{\parallel}$. For AlN we find $d\ln m^*/d\ln V$ is $-1.1$ for $m_e^{\perp}$ and $-1.2$ for $m_e^{\parallel}$. For the case of biaxial stress, tensile strain causes $m_e^{\perp}$ in GaN and AlN to decrease at a similar rate as in the case of hydrostatic stress, while $m_e^{\parallel}$ *increases* with tensile strain.

### Implications for $\mathbf{k}\cdot\mathbf{p}$

We examine this unexpected variation of $m_e^{\parallel}$ with strain by employing the $\mathbf{k}\cdot\mathbf{p}$ formalism of Section 2.2.[59,60] Under the quasi-cubic approximation and neglecting the effects of other bands and spin-orbit coupling, the electron effective mass in the $\mathbf{k}\cdot\mathbf{p}$ formalism can be expressed as:[59]

$$\frac{m_0}{m_e^{\parallel}} \simeq 1 + \frac{E_{\parallel}}{E_g + \Delta_{cr}}, \tag{2.13}$$

and

$$\frac{m_0}{m_e^{\perp}} \simeq 1 + \frac{E_{\perp}}{E_g}, \tag{2.14}$$

where $E_{\parallel}$ and $E_{\perp}$ are the so-called Kane parameters. These are related to the interband momentum matrix elements as follows: $E_{\parallel} = \frac{2}{m_0}|\langle CF|\hat{p}_{\parallel}|CB\rangle|^2$ and $E_{\perp} = \frac{2}{m_0}|\langle HH|\hat{p}_{\perp}|CB\rangle|^2 = \frac{2}{m_0}|\langle LH|\hat{p}_{\perp}|CB\rangle|^2$, where $CB$ denotes the conduction-band wave function.[58] This model is usually applied by fitting the Kane parameters to experimental

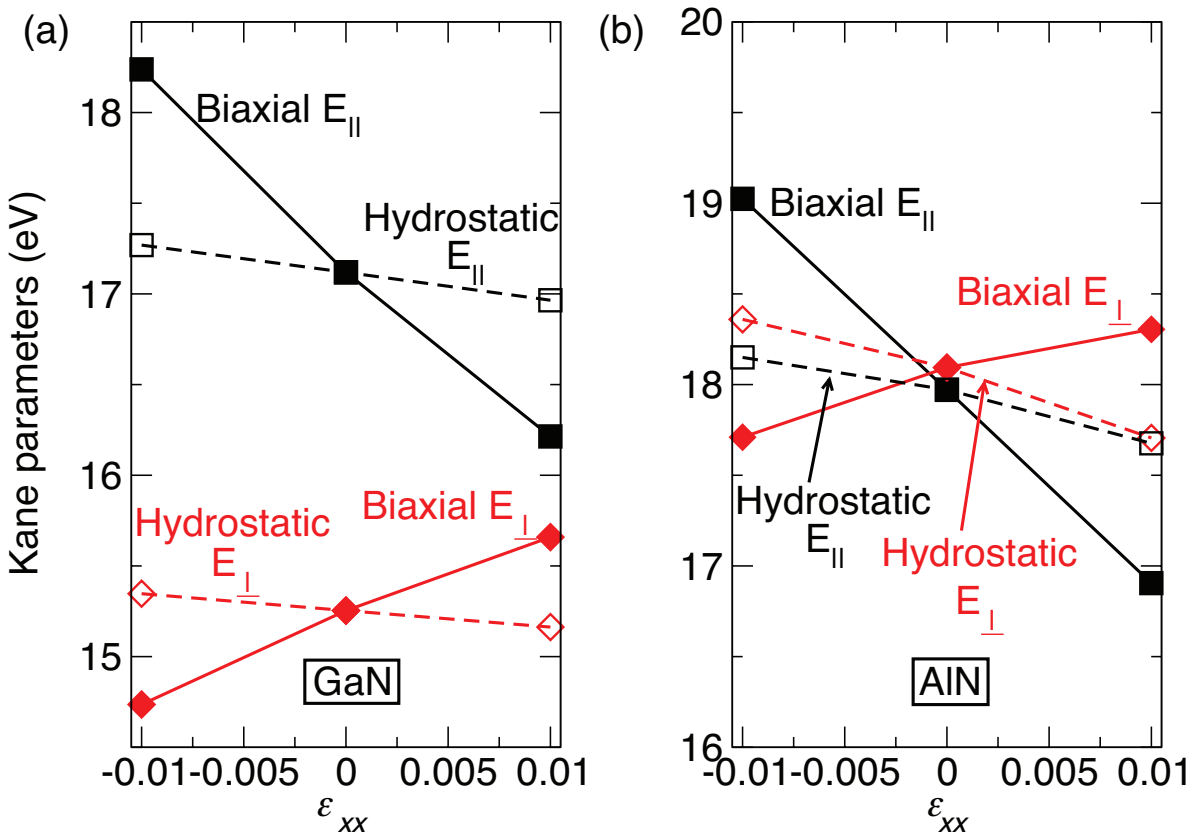


**Figure 2.4:** Dependence of Kane parameters on strain for (a) GaN and (b) AlN under hydrostatic and biaxial stress.

or calculated data. Past applications of the model have assumed that the Kane parameters are constant, i.e., that they do not depend on strain. Here we find that there is a distinct variation with strain, and we explicitly determine the parameters by calculating the momentum matrix elements between the wave functions for the conduction band and the three highest valence bands at different strains. The results are shown in Fig. 2.4.

In the case of hydrostatic stress we find that the Kane parameters vary only slightly as a function of strain; therefore, according to Eqs. (2.13) and (2.14), the change in effective mass is dominated by the change in $E_g$, leading to the often-invoked trend that effective mass decreases with decreasing band gap (and thus with increasing hydrostatic stress, Fig. 2.3). In the case of biaxial stress, however, the Kane parameters change significantly with strain, with $E_{\parallel}$ and $E_{\perp}$ exhibiting opposite trends. This explains the opposite behavior in $m_e^{\parallel}$ and $m_e^{\perp}$ under biaxial stress, observed in Fig. 2.3. Therefore, for $\mathbf{k} \cdot \mathbf{p}$ to correctly capture the strain dependence of the effective mass in GaN and AlN, the effect of strain on the momentum matrix elements must be taken into account.

We can use our HSE-calculated gaps and strain-dependent Kane parameters in the $\mathbf{k} \cdot \mathbf{p}$ formalism outlined in Section 2.2, and compare the strain dependence with direct fitting of the HSE dispersion shown in Fig. 2.3. We see that the trends with strain in Fig. 2.3 are captured correctly in Fig. 2.5 , though the absolute values of the effective masses and the anisotropy in effective mass are underestimated. To explain this, we consider the effects of additional bands beyond the 8x8 formalism (Section 2.2) within second-order nondegenerate perturbation theory. The interaction between the conduction band with other bands will change the conduction band's (inverse) effective mass by a factor[57]

$$\Delta \frac{m_0}{m_e^*} = \frac{|p_{mn}|^2}{E_n(0) - E_m(0)}, \tag{2.15}$$

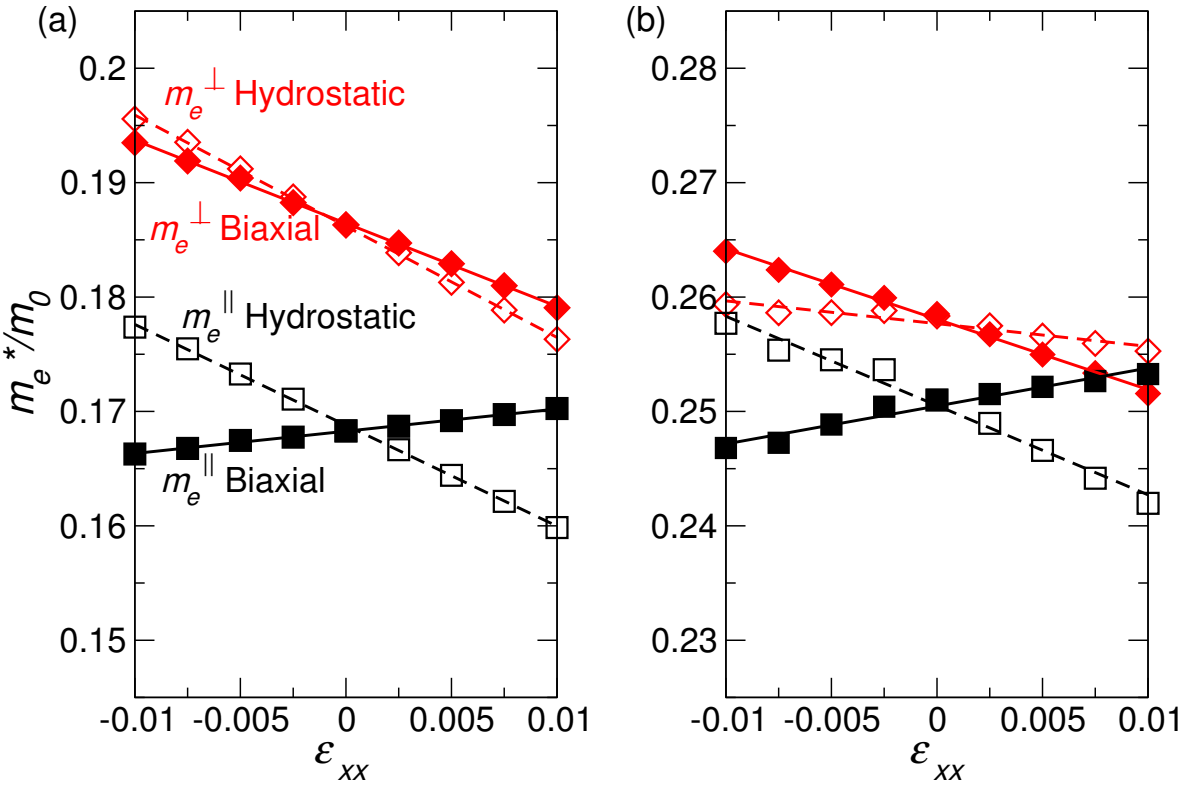


**Figure 2.5:** Dependence of effective mass on in-plane strain according to $\mathbf{k} \cdot \mathbf{p}$ theory with the gaps and strain-dependent effective masses taken from experiment. (a) GaN and (b) AlN, for hydrostatic as well as biaxial stress. The parallel ($m^{\parallel}$) and in-plane ($m^{\perp}$) directions are indicated in Fig. 2.1.

where the subscript $n$ denotes the conduction band, and $E(0)$ denotes the energy at the Γ-point. Therefore, interactions with additional valence bands will result in a positive $\Delta\frac{m_0}{m_e^*}$, and therefore a decrease in $\frac{m_e^*}{m_0}$. We find that the lower-lying valence bands have negligible matrix elements and can be neglected. Interactions with additional conduction bands results in a negative $\Delta\frac{m_0}{m_e^*}$, which would increase $\frac{m_e^*}{m_0}$. The matrix elements between the lowest conduction band and higher lying empty bands can be significant, and neglecting their effect causes the underestimation of magnitudes of the effective masses by 8x8 $\mathbf{k}\cdot\mathbf{p}$ theory.

**Implications for HEMT devices**

The calculated change in effective mass as a function of strain can be used to estimate the effect of strain on the electron mobility. For parabolic bands in the semiclassical model of conduction, the in-plane conductivity is given by[68]

$$\sigma_\perp = \frac{ne^2\tau}{m_e^\perp}, \tag{2.16}$$

where $n$ is the number of carriers per unit volume and $\tau$ is the relaxation time, or the average time between scattering events. The electron mobility is given by the relaxation time divided by the effective mass. The in-plane mass, $m_e^\perp$, decreases with tensile strain at a rate of around 5 % for a strain of $\epsilon_{xx} = 0.01$. If all other variables in Eq. (2.16) are kept constant, this corresponds to an increase in mobility of less than 1%, over an

order of magnitude too small to explain the increase in mobility reported in Ref. 55 based on measurements of strained HEMTs. The variation in effective mass with strain is therefore clearly not the main contributor to the reported change in mobility. Such an enhancement, therefore, would require a substantial change in $\tau$ due to strain effects on scattering mechanisms in the two-dimensional electron gas. These effects are beyond the scope of the present study.

**Hole effective masses**

For GaN, the determination of the hole effective masses from the HSE dispersions requires extra care since the three highest valence bands are close to each other in energy, and the character of the highest valence band changes with strain. Therefore, a very dense mesh of $k$-points near the crossings is necessary to distinguish between bands and provide accurate values of the curvature, or some type of band decomposition (based on Wannier functions) is needed. Here we present the influence of strain on the in-plane masses for the valence bands of GaN for reference. For the reasons just mentioned, there is considerable scatter in the data, but an overall downward trend with tensile strain is discernible.

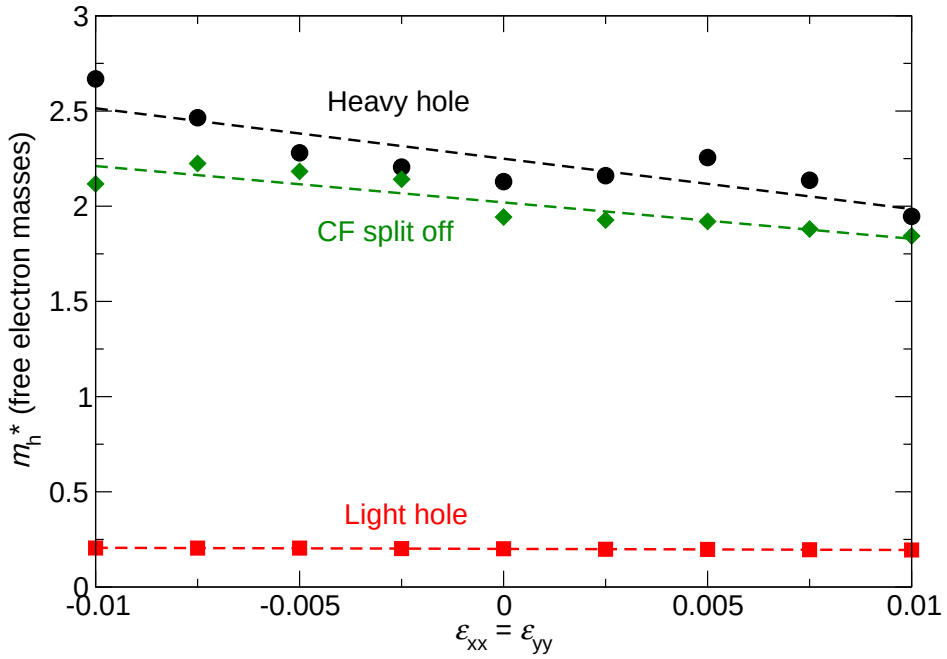


**Figure 2.6:** Dependence of $m^{\perp}$ hole effective mass on biaxial strain for the three highest valence bands in GaN.

## 2.4 Conclusions

We have systematically studied the effect of strain on the effective mass of AlN and GaN using first-principles techniques based on hybrid density functional theory. Hydrostatic tensile stress causes a linear decrease of both $m_e^{\perp}$ and $m_e^{\parallel}$. Biaxial stress also results in a linearly decreasing $m_e^{\perp}$, but $m_e^{\parallel}$ increases linearly. An examination of the 8×8 $\mathbf{k} \cdot \mathbf{p}$ formalism shows that this behavior is caused by the strain dependence of the momentum matrix element between the valence and conduction bands, which is substantial for biaxial stress states. Our results show that the change of effective mass with stress cannot account for the changes in mobility reported in Ref. 55.

# Chapter 3

# Surface energies and cracking in III-nitrides

The surface energies of different planes in a crystal determine the equilibrium morphology[69] and equilibrium growth rate in the direction normal to each plane.[70] Absolute surface energies are therefore key to understanding and optimizing growth processes of bulk or epitaxial films. In the case of III-nitrides, the lack of availability of bulk substrates currently requires performing heteroepitaxial growth, affecting the crystal quality and sometimes leading to facet formation or to the presence of residual strains that cause cracking, another effect which is determined by surface energies. Processes such as selective area growth (SAG) are used to improve the quality of heteroepitaxial films,[71] hydride vapor phase epitaxy (HVPE) is employed to produce thick substrate-like films,[72] and

actual bulk growth is being explored using techniques such as ammonothermal growth[73] or Na flux.[27] All of these would benefit from having accurate values for surface energies.

Here we present first-principles calculations of absolute surface energies in GaN [Section 3.1, reprinted excerpt with permission from Ref. 29, copyright (2014) American Physical Society], and their application to critical thicknesses for cracking of AlGaN on GaN (Section 3.2).

## 3.1 Absolute surface energies of GaN from first principles

### 3.1.1 Overcoming computational and fundamental challenges

The fact that accurate values for nitride surface energies are not yet available is due to two issues: (1) absolute energies for polar or semipolar planes are fundamentally ill-defined for a crystal with low symmetry such as wurtzite (wz), as shown rigorously in Ref. 74, and further discussed in Ref. 75; and (2) uncertainties in previously calculated results due to limitations in the computational techniques. In this work, we have overcome these problems and calculated the absolute surface energies of the polar and nonpolar planes in GaN from first principles; we have investigated a large number of reconstructions for bare as well as hydrogen-covered surfaces to connect to a vari-

ety of conditions representing experimental growth by molecular beam epitaxy (MBE), metal-organic chemical vapor deposition (MOCVD), and HVPE.

The indeterminacy of the individual surface energies for the polar surfaces is a fundamental problem that is intrinsic to the wurtzite crystal structure. Here we approach the issue pragmatically by invoking the similarity between polar $(0001)/(000\bar{1})$ planes in wz and $(111)/(\bar{1}\bar{1}\bar{1})$ planes in zincblende (zb), for which absolute surface energies *can* be determined due to the higher symmetry of the zb crystal structure (see Fig. 3.1).[74,76] Such an approach was previously applied to CdSe.[77,78] The second issue, accuracy of the computational results, is addressed by our use of a hybrid functional in the first-principles calculations based on density functional theory (DFT). This approach overcomes the deficiencies of traditional functionals such as the local density approximation (LDA) or generalized gradient approximation (GGA), which severely underestimate the band gap (by as much as 50% in GaN). The band-gap error also affects the energies of surface states within the gap,[79] and when these states are occupied with electrons (which is invariably the case for all relevant surface reconstructions) the calculated surface energy is also affected.

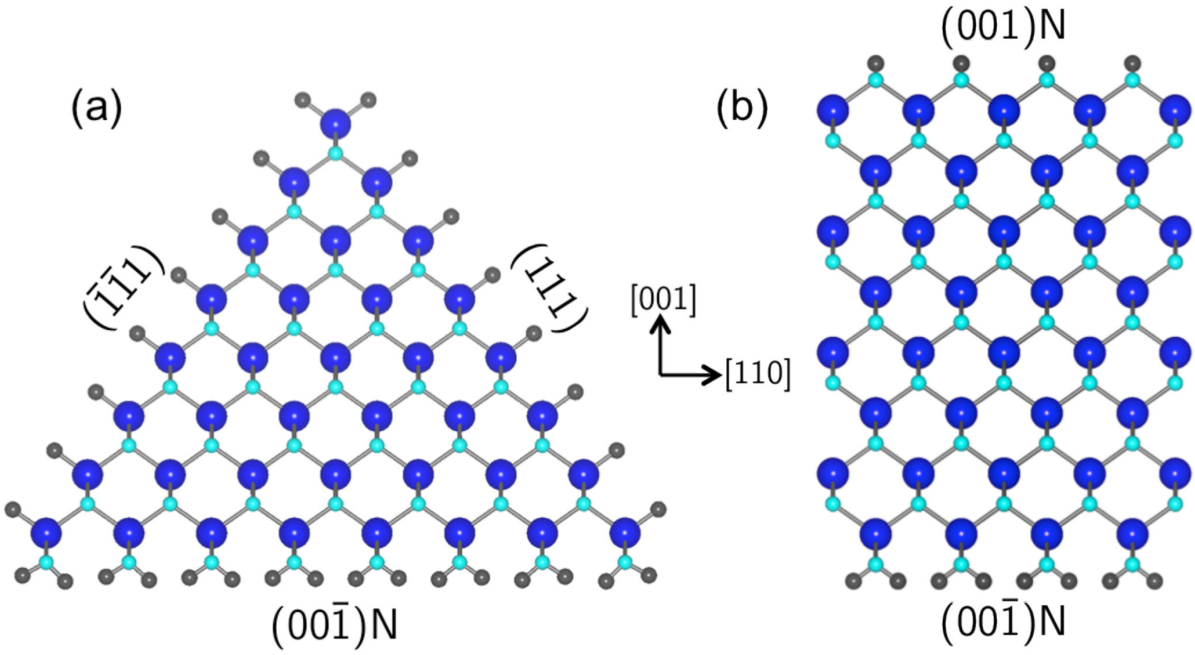


**Figure 3.1:** (a) Cross section of a zb-GaN triangular wedge used to determine the energy of the (111) surface. The bottom surface is a N-terminated (001)N surface, the other surfaces are symmetrically equivalent (111) and ($\bar{1}\bar{1}1$) surfaces. All surfaces are passivated with fractional H atoms. (b) Cross section of the zb-GaN slab used to determine the energy of the passivated {001} surface.

### 3.1.2 Methodology to obtain separate polar surface energies: The zincblende wedge cell method

Absolute surface energies of the nonpolar $\{10\bar{1}0\}$ $m$ and $\{11\bar{2}0\}$ $a$ planes are rigorously defined and accessible from slab calculations that can be performed within the supercell approach.[80] For a supercell containing a slab with $n$ formula units, separated by an appropriate amount of vacuum, the surface energy is obtained by subtracting

the total energy of $n$ formula units of bulk GaN from the total energy of the slab and dividing by twice the total surface area (since the cell contains two identical surfaces).

GaN slabs terminated by polar planes inevitably exhibit inequivalent surfaces; a surface energy obtained from a slab calculation therefore yields the *average* surface energy of the (0001) ($+c$) and (000$\bar{1}$) ($-c$) planes. In addition, spurious charge transfer from the Ga dangling bonds on the $+c$ to the N dangling bonds on the $-c$ surface is likely to occur; this problem can be avoided by passivating the bottom surface of the slab with fractionally charged hydrogen. By keeping the structure of the bottom surface fixed, energy differences can be determined between various reconstructions of the top surface for a particular polar plane.[80–84] This procedure yields relative energies; the absolute surface energy of the top surface can still not be determined without knowledge of the absolute surface energy of the passivated bottom surface.

In this work, we assume the energies of the wz passivated bottom surfaces are equal to the energies of passivated zb (111) or ($\bar{1}\bar{1}\bar{1}$) planes. The zb phase of GaN is energetically similar to wz (we calculate an energy difference of 26 meV per formula unit), and the atoms on the zb (111)/($\bar{1}\bar{1}\bar{1}$) planes have the same coordination and structure up to the next-nearest neighbors of the surface atoms as the wz polar planes. In addition, the calculated sum of the surface energies of the passivated (0001) and (000$\bar{1}$) planes (a well defined quantity) differs from the sum of the (111) and ($\bar{1}\bar{1}\bar{1}$) surface energies by less than 2 %. At first sight, the zb (111)/($\bar{1}\bar{1}\bar{1}$) planes may seem to suffer from the same

problem as wz polar surfaces in that a slab oriented along the [111] direction exhibits inequivalent top and bottom surfaces. However, the higher symmetry of the zb phase (in particular, the existence of two rotation axes[74]) ensures that a rigorous definition of the absolute (111)/($\bar{1}\bar{1}\bar{1}$) surface energy *is* possible,[75] and a practical method was proposed by Zhang and Wei[76] based on calculations for an infinitely long wedge with triangular cross section [Fig. 3.1(a)].

The wedge surfaces are all passivated with fractional hydrogen atoms (i.e., hydrogen-like atoms with an electron/proton charge of $0.75e$ for the N-terminated face and $1.25e$ for the Ga-terminated face). A similar wedge can be constructed with ($\bar{1}\bar{1}\bar{1}$) surfaces and a Ga-terminated (001) surface.

**Isolating the surface energy of the polar zincblende {111} planes**

There are four contributions to the total energy of a such a "wedge" cell containing $n$ formula units of GaN:

$$E_{wedge}^{(111)}(n) = E_{bulk}(n) + A^{(111)}\sigma_{pass}^{(111)} + A^{(001)}\sigma_{pass}^{(001)} + \sum E_{edges}, \tag{3.1}$$

where the first term is the energy associated with a corresponding number of bulk GaN formula units, the second term is the contribution from the surface energy of the (111) surface, the third term is the surface energy of the {001} surface, and the last term is the contribution from the edges of the wedge.

Assuming that the contribution to the energy from the edges is the same for different size cells (if the cells are sufficiently large), this contribution can be cancelled by subtracting the total energies of wedges of different sizes (here we use $n = 36$ and 28). The resulting expression for a cell with (111) surfaces (Fig. 3.1) is then

$$\begin{aligned} E^{(111)}_{wedge}(n=36) - E^{(111)}_{wedge}(n=28) - E_{bulk}(n=8) \\ = 2A^{(111)}\sigma^{(111)}_{pass} + A^{(001)}\sigma^{(001)N}_{pass} \end{aligned} \tag{3.2}$$

And the analogous expression for a cell with $(11\bar{1})$ surfaces (made by replacing the Ga with N and vice versa in Fig. 3.1) is

$$\begin{aligned} E^{(\bar{1}11)}_{wedge}(n=36) - E^{(\bar{1}11)}_{wedge}(n=28) - E_{bulk}(n=8) \\ = 2A^{(111)}\sigma^{(\bar{1}11)}_{pass} + A^{(001)}\sigma^{(001)Ga}_{pass} \end{aligned} \tag{3.3}$$

The contribution from the $\{001\}$ surface can be determined in a straightforward way by using a slab cell with two equivalent (Ga or N terminated) surfaces. These cells, however, are not stoichiometric, and therefore, the additional energy required to add an extra Ga or N surface atom must be accounted for by reference to the species' chemical potentials. If we assume that the GaN bulk is in thermodynamic equilibrium with a reservoir of Ga atoms (Ga bulk) and N atoms (gas of $N_2$ molecules), then the chemical potentials of Ga and N are related through the expression

$$\mu_{Ga} + \mu_N = \mu_{GaN} = \mu_{Ga(bulk)} + \frac{1}{2}\mu_{N_2} + \Delta H_f, \tag{3.4}$$

where $\Delta H_f$ is the enthalpy of formation of GaN, which we calculate to be $-1.3$ eV. We reference all of the chemical potentials to the energy per atom of bulk Ga, $\mu_{Ga(bulk)}$. The chemical potential of Ga is therefore limited to the range from 0 eV (equal to $\mu_{Ga(bulk)}$, Ga rich conditions) to $-1.3$ eV (N chemical potential equal to $\frac{1}{2}\mu_{N_2}$, N rich conditions).

Rempel *et al.*[78] subsequently showed that better convergence can be obtained if an energy difference is considered between structures of the same size in which cations and anions are interchanged, i.e., wedges terminated by (111) planes versus $(\bar{1}\bar{1}\bar{1})$ planes. Altogether, the difference in surface energy $\sigma_{\rm pass}^{(111)} - \sigma_{\rm pass}^{(\bar{1}\bar{1}\bar{1})}$ is given by

$$\begin{aligned}\sigma_{\rm pass}^{(111)} - \sigma_{\rm pass}^{(\bar{1}\bar{1}\bar{1})} = \frac{1}{2A^{(111)}}\Big\{\big[E_{\rm wedge}^{(111)}(36) - E_{\rm wedge}^{(\bar{1}\bar{1}\bar{1})}(36)\big] - \big[E_{\rm wedge}^{(111)}(28) - E_{\rm wedge}^{(\bar{1}\bar{1}\bar{1})}(28)\big] \\ -\frac{1}{2}\big[E_{\rm slab}^{(001){\rm N}} - E_{\rm slab}^{(001){\rm Ga}}\big] + \frac{1}{2}\big[\mu_{\rm N} - \mu_{\rm Ga}\big]\Big\}\end{aligned} \tag{3.5}$$

where $E_{\rm wedge}^{(111)}(n)$ is the total energy of the wedge [Fig. 3.1(a)] with passivated (111) surfaces containing $n$ GaN formula units and $E_{\rm slab}^{(001){\rm Ga/N}}$ is the total energy of a slab with (001)Ga/N surfaces [Fig. 3.1(b)]. As noted above, the *sum* of the polar surface energies, $\sigma_{\rm pass}^{(111)} + \sigma_{\rm pass}^{(\bar{1}\bar{1}\bar{1})}$ is easily obtained from a slab calculation, and combining that value with Eq. (3.5) then gives the individual absolute energies for the passivated surfaces.

The absolute surface energy for a reconstructed wz polar surface is then given by

$$\sigma^{(0001)} = \frac{1}{A^{(0001)}}\big(E^{(0001)}(n) - E_{\rm bulk}(n) - n_{\rm Ga}\mu_{\rm Ga} - n_{\rm N}\mu_{\rm N} - A^{(0001)}\sigma_{\rm pass}^{(\bar{1}\bar{1}\bar{1})}\big), \tag{3.6}$$

where $E^{(0001)}$ is the total energy of a slab with the $-c$ face passivated and the $+c$ face reconstructed and $n_{\mathrm{Ga}}/n_{\mathrm{N}}$ is the number of atoms added or removed to form the reconstructed surface.

**Computational parameters**

The total energy calculations were performed using the HSE hybrid functional[43,46] with 31 % mixing of Hartree-Fock exchange, and projector-augmented-wave[50] pseudopotentials as implemented in the VASP[49] code. The energy cutoff of the plane-wave basis set was set to 300 eV, with convergence checks up to 400 eV.

The wedge cells [Fig. 3.1(a)] contained 28 or 36 formula units of zb-GaN separated by at least 15 Å of vacuum. A 1x1x4 Monkhorst-Pack[53] $k$-point mesh was used for integrating over the Brillouin zone (4 $k$-points in the direction along the wedge axis). Keeping the geometry of the passivated $(111)/(\bar{1}\bar{1}\bar{1})$ surfaces identical between the wedge and slab cells is essential for consistency. The fractional hydrogen atoms on the wedge surfaces were first allowed to relax with the Ga and N atomic positions fixed. The H atoms along with the surface layer of GaN were then fixed, and the bulk-coordinated GaN atoms were allowed to relax. These fixed surface geometries were used for the first layer of the wz-GaN passivated bottom surfaces for the calculation of the reconstructions on the $+c$ and $-c$ planes.

The zb {001} surface-energy slab cells contained 8 bilayers of GaN, and 18 Å of vacuum; the reconstructed wz slabs were 2x2 or $\sqrt{3}$x$\sqrt{3}$ unit cells by 12 bilayers of GaN and 18 Å of vacuum; the *a* plane slab cells contained 14 bilayers of GaN and 12 Å of vacuum; the *m* plane slab cells contained 12 bilayers of GaN and 12 Å of vacuum. A 4x4x1 *k*-point mesh was used for all slab calculations. We estimate the numerical convergence of surface energies to be within 0.02 eV/Å$^2$.

### 3.1.3 Absolute surface energies of GaN

The results for bare surfaces are presented in Fig. 3.2. For clarity, only the lowest-energy reconstruction for a specific surface plane at each value of $\mu_{\mathrm{Ga}}-\mu_{\mathrm{Ga(bulk)}}$ is plotted in Fig. 3.2. Those lowest-energy reconstructions for given planes are the same as those found in previous calculations,[81,84] but our results now allow comparing *absolute* energies for different planes. Our calculated surface energies for the unreconstructed (but of course relaxed) nonpolar planes agree within 9 % with values obtained using DFT-LDA.[85] For polar surfaces, we have found sizeable differences between absolute surface energies calculated with HSE and less accurate functionals such as GGA, confirming the importance of the use of a hybrid functional.

In order to address MOCVD, HVPE or ammonia MBE growth, we need to study reconstructions involving hydrogen and explicitly consider *free* energies, both because of the temperature and pressure dependence of the chemical potential of gaseous sources

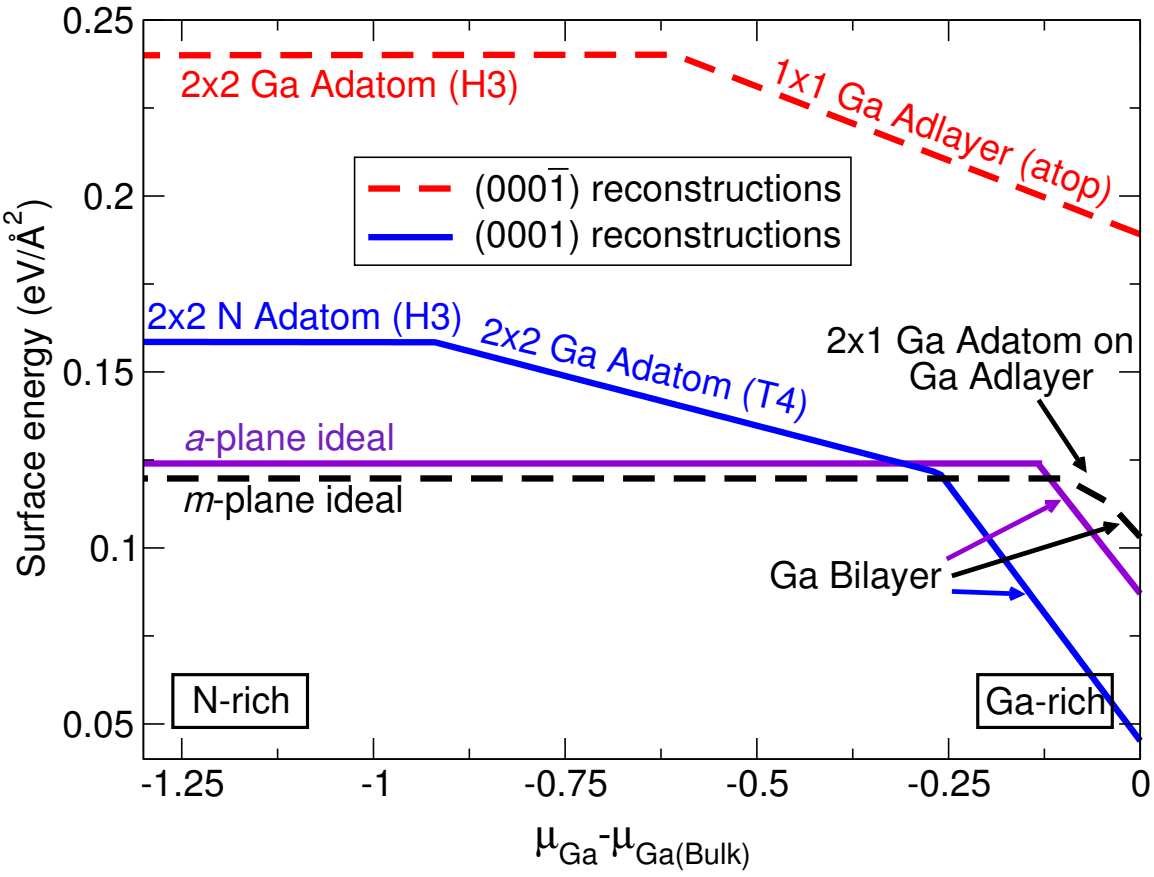


**Figure 3.2:** Absolute surface energies of GaN nonpolar and polar planes as a function of Ga chemical potential.

and because of entropy contributions from vibrational modes of adsorbates, which can be sizeable due to the small atomic mass of hydrogen;[86,87] the latter have been determined as in Ref. 87.

Figure 3.3 shows the lowest-energy hydrogenated polar surface reconstructions for $T$=1300 K, $p$=1 atm, representative of MOCVD growth conditions, and $T$=1000 K, $p$=$10^{-12}$ atm, representative of MBE growth.[88] The effects of temperature and pressure conditions that differ from these specific choices can be estimated by considering that increasing $T$ increases the surface energy of the hydrogenated reconstructions (mainly due to a decrease in $\mu_{\mathrm{H}}$), while increasing $p$ (increasing $\mu_{\mathrm{H}}$) decreases the surface energy.

The energies in Fig. 3.3 can be compared with those of the bare surfaces in Fig. 3.2 (free-energy contributions have only minor effects on the latter). We find that hydrogenated reconstructions significantly decrease the surface energies of the $+c$ and $-c$ planes, except under very Ga-rich conditions; reconstructions are designated by the same labels as in Refs. 87 and 88. On the other hand, reconstructing the nonpolar surfaces with hydrogen did not lower the surface energy.[86] The similarity in surface energies between the $m$ and $a$ planes over a wide range of conditions is consistent with observations of similar growth rates of the planes in MOCVD and HVPE SAG.[71,72,89–91]

Even when considering H-containing reconstructions, the $+c$ surface remains lower in energy than the $-c$ over the whole Ga chemical potential range, though this difference is significantly reduced under N-rich conditions at 1300 K, 1 atm. On the other hand,

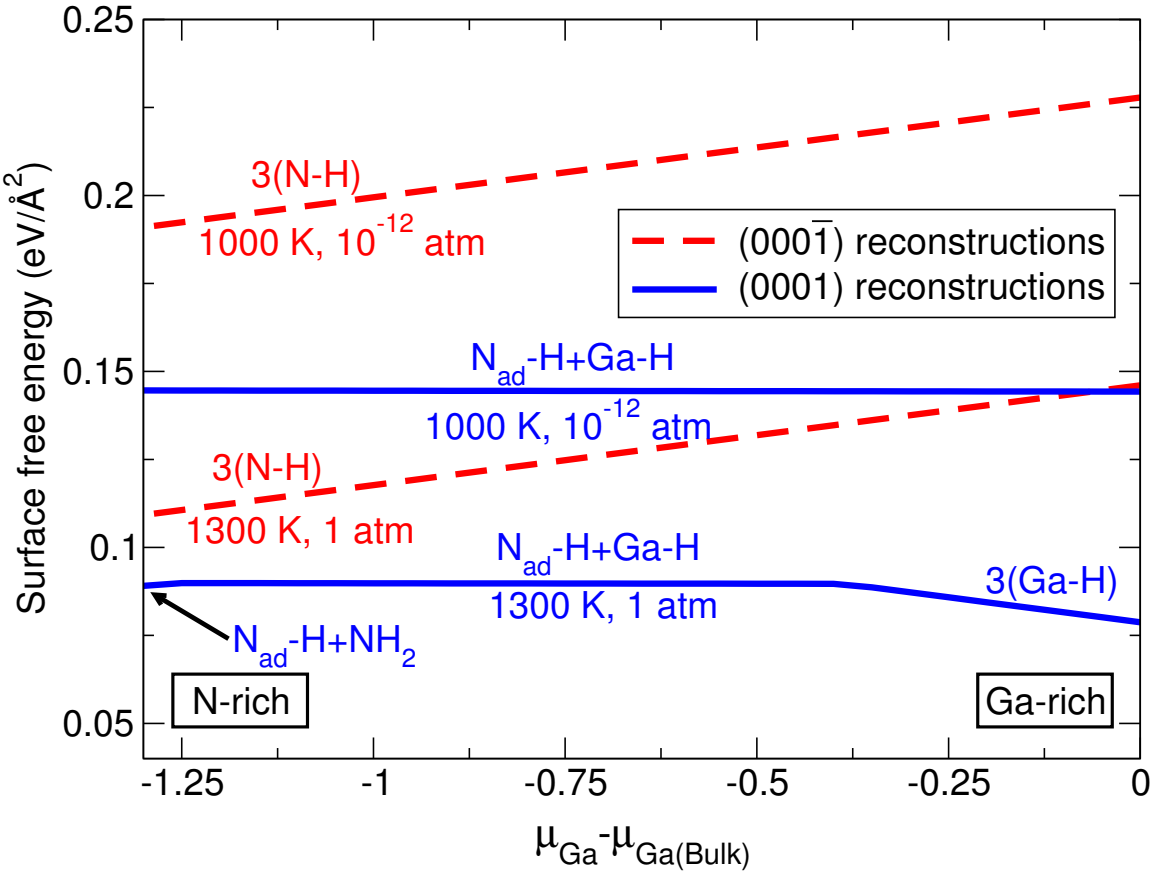


**Figure 3.3:** Absolute surface free energies of hydrogenated GaN polar planes for different temperature and pressure conditions, as a function of Ga chemical potential. Hydrogenation of nonpolar surfaces does not lower the surface energy at these $T$ and $p$ conditions.

hydrogenation of the polar planes significantly alters their relative stability with respect to the nonpolar planes (which remain bare). For MOCVD conditions, the hydrogenated $+c$ surface is lower in energy than the nonpolar planes over the entire chemical potential range, and the $-c$ plane is more stable than the nonpolar planes under N-rich conditions.

In SAG experiments, the $-c$ plane is observed to grow much more slowly than the $+c$,[71,72,89] contrary to the equilibrium growth we would expect from our surface energies. The difference could be due to kinetic factors, or possibly due to a lower-energy surface reconstruction on the $-c$ surface that has not yet been considered.

### 3.1.4 Conclusions

In summary, we have calculated absolute surface energies for nonpolar and polar orientations of GaN, for bare and hydrogenated surfaces. The nonpolar $a$ and $m$ planes are similar in energy, and their energies do not depend on the growth conditions over a wide range of temperatures, pressures, and chemical potentials. The $+c$ face is lower in energy than the $-c$ face for all conditions considered. Hydrogenation of the polar surfaces under relevant growth conditions significantly reduces their energy, rendering them more stable than the nonpolar planes. These results will form a sound basis for analyzing and guiding growth of GaN with a variety of techniques, as well as providing a framework to determine surface energies of other technologically relevant wurtzite materials such as InN, AlN and ZnO.

## 3.2 Cracking in nitrides

As mentioned previously, due to the lack of widely available bulk substrates and lattice mismatch between layers, residual stresses are invariably present in nitride heterostructure devices. This is particularly problematic when the application requires thick layers, such as for cladding layers in lasers and barriers to confine the two-dimensional electron gas in high electron mobility transistors. As the film grows thicker, the strain energy increases until a certain "critical thickness," where stress relaxation occurs, resulting in a degradation in structural quality. Therefore, in order to optimize the performance of devices, it is necessary to understand the mechanisms that govern this relaxation.

The critical thickness and the strain relaxation mechanism depend on the growth plane of the film;[92,93] while most nitride films are still grown in the polar (0001) or $(000\bar{1})$ $c$ plane, growth on the nonpolar and semipolar planes can improve the properties of optoelectronic devices.[94] Here we consider the critical thickness for growth on the $c$ plane and the nonpolar $\{10\bar{1}0\}$ $m$-and $\{11\bar{2}0\}$ $a$ planes.

The relaxation mechanism also depends on the type of strain in the film, which varies depending on the substrate and heterostructure material. Compressive strain is present in InN or InGaN films grown on GaN, or is due to lattice and thermal-expansion-coefficient mismatch for GaN on sapphire. This strain can be relaxed through film delamination, and will not be considered here. Tensile strain is present in $Al_xGa_{1-x}N$ grown

on GaN and GaN or $Al_xGa_{1-x}N$ films grown on Si substrates due to the lattice mismatch, and GaN or $Al_xGa_{1-x}N$ grown on SiC due to the thermal-expansion-coefficient mismatch. Also, tensile strain can be present in GaN films grown on sapphire caused by island coalescence.[22] Tensile strain can be relaxed by cracking.[23]

In principle, both tensile and compressive strain can also be relaxed plastically, by the formation of misfit dislocations (MDs). However, the wurtzite (wz) structure of III-nitride films limits the efficacy of plastic relaxation. In a strained film, MDs form as a result of the glide of threading dislocations (TDs) when the shear stress resolved on a glide plane is substantial enough to overcome the energetic penalty to form additional dislocation length (line tension of the TD)[95] and the Peierls force[96] for the glide plane. For the wz structure, the most favorable glide plane is the $c$ plane (basal slip); however, for films grown in the $c$ and nonpolar planes, there is no shear stress resolved on the $c$ plane, so basal slip cannot occur.[93] Prismatic glide on the $m$ planes has been observed,[97] though is expected to have large Peierls forces.[98] In addition, it is not relevant for $c$ plane growth since it does not result in strain relaxation parallel to the $c$ plane.[98] Therefore, only the pyramidal glide planes ($\{11\bar{2}2\}$, $\{1\bar{1}02\}$, or $\{10\bar{1}1\}$), which also are expected to have high Peierls forces,[98] can result in strain relaxation in $c$ plane films.

Because of these limitations on plastic deformation, the initial stress relaxation for films under tensile strain grown on the $m$- and $c$ plane occurs through the formation of cracks.[22,92,99–108] Once cracks form, they can create local shear stresses and nucleation

sites for the creation of MDs, further reducing stress.[99,100,102,103] Therefore, the relevant critical thickness for tensile strain relaxation in nitrides is that for the formation of cracks.

From the Griffiths criterion, the critical thickness, $h_c$, for a crack to propagate in a isotropic medium under equibiaxial loading is given by[23]

$$h_c = \frac{\Gamma \bar{E}}{Z\sigma^2} \tag{3.7}$$

where $\sigma$ is the tensile stress resolved on a given plane, $\bar{E}$ is the biaxial elastic modulus, and $\Gamma$ is the fracture toughness of the material, discussed below.

In Eq. (3.7), $Z$ is a dimensionless factor that describes the geometry of the crack. In strained epitaxial films, small, isolated, penny-shaped "surface cracks" initially form at flaws. These cracks require a relatively small amount of energy to develop ($Z$=3.951, Ref. 23) and are stable as unconnected slits. As strain energy increases with increasing thickness, the "channeling" process becomes activated, during which the surface cracks extend in length and depth until they terminate at the substrate, a free surface, and/or another channel crack. For this process, $Z$=1.976,[23] indicating that it requires more energy to activate.

Equation (3.7) reflects an energetic balance between strain energy density in the film ($h\sigma^2/\bar{E}$) and the energy increase associated with a propagating crack, or the fracture toughness, $\Gamma$. In general, $\Gamma$ contains contributions due to the surface-energy increase

when new crack-face area is created as well as the energy of plastic deformation at the crack tip;[23] such plastic deformation increases the energy necessary to propagate the crack, so it will result in a higher fracture toughness, and therefore a larger critical thickness. Here we assume brittle fracture; the fracture toughness in the brittle limit, $\Gamma_b$, is simply given by the surface-energy increase, providing a lower bound on the true fracture toughness.

For growth on the $c$ plane, cracks have been observed exclusively on the nonpolar planes,[22,99–102] while for nonpolar growth, cracks are observed on the $c$ plane.[92] Therefore, we have calculated $\Gamma_\mathrm{b}$ for cracks to form on the nonpolar $a$ and $m$ planes, as well as the polar $c$ planes in wurtzite GaN and AlN from first principles.

### 3.2.1 First-principles calculations of surface energies relevant for cracking in AlN and GaN

The first step to calculating surface energies relevant to cracking is to determine the morphology that is likely to be present on a given plane that is created as a result of cleaving. As discussed in Section 3.1, atomic reconstructions on the polar and nonpolar planes have been shown to reduce the surface energies in GaN and AlN over the ideal, unreconstructed surface for certain conditions,[84,86,109,110] which is especially important for the polar planes, where the surface-energy reduction is significant for all conditions

under which GaN (AlN) is stable. Therefore unreconstructed surfaces are unlikely to be present when the crystal cleaves and do not represent the surface-energy increase as the crack propagates.

We take into account relevant reconstructions with a very simple model of cracking in these materials: we assume that since the cracks propagate at speeds on the order of the Rayleigh wave speeds ($\sim$ km/s),[111] there is not enough time for atoms (either excess Ga or N, or impurity atoms such as H) to diffuse to the crack tip either through the void created by the crack, or along the surface of the crack face. Therefore, a given surface formed by the crack can only reconstruct using atoms taken from the opposite surface, or nearby (within one or two unit cells) on its own surface (see Fig. 3.4). In other words, we have conserved the number of atoms between the two surfaces in a simulation cell with a specific crack area (2×2 or $\sqrt{3} \times \sqrt{3}$ unit cells, the size of the unit cell of typical reconstructions in nitrides[84]). This effectively excludes reconstructions with Ga adlayers that were included in Section 3.1; such reconstructions have been found to be stable under some growth conditions, but no reservoir of Ga atoms is available during crack formation. To obtain the fracture toughness, we choose the lowest-energy reconstructed surfaces under these stoichiometry-preserving constraints, which gives a lower bound on the surface-energy increase as the crack propagates.

For the nonpolar planes, the surface energies are determined from slab cell calculations as in the case of the absolute surface energy calculations discussed in Section

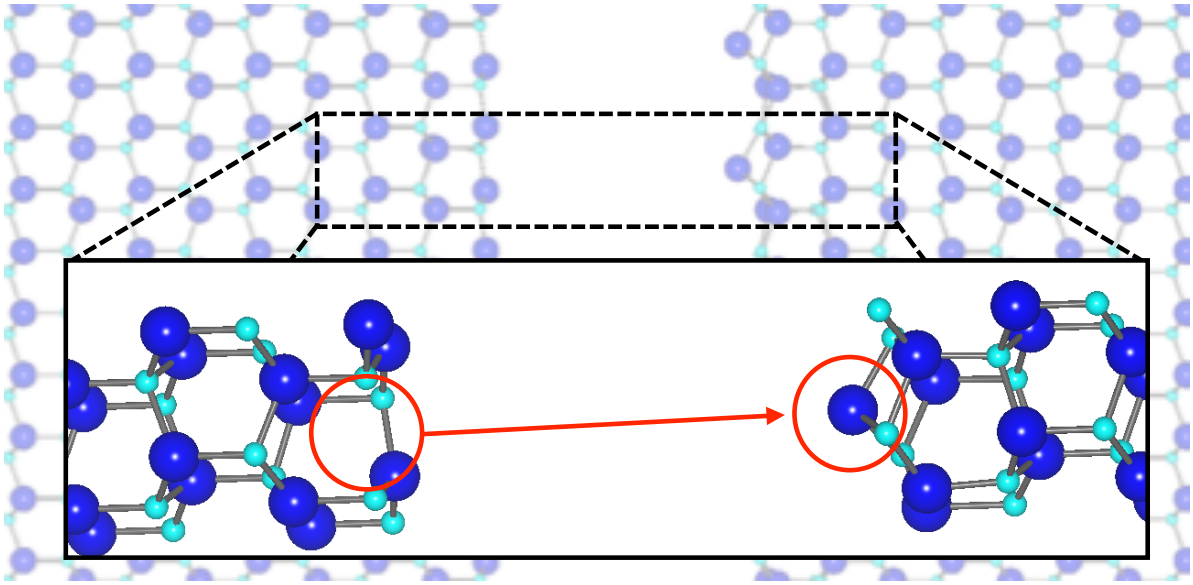

**Figure 3.4:** Enlargement of the local structure of a $c$ plane crack, with a stoichiometry-preserving reconstruction.

3.1. To obtain brittle rapture toughnesses, $\Gamma_{\mathrm{b}}^{\mathrm{np}}$, we simply have to multiply the absolute surface energies by two (since a crack forms two free surfaces).

For a slab oriented along the $c$ direction, both $+c$ (0001) and $-c$ (000$\bar{1}$) are necessarily present. This caused serious problems for absolute surface energies (Section 3.1), but actually corresponds to the physical situation of a cleavage in the crystal, where there will necessarily have to be a $+c$- and $-c$-derived surface as the crack propagates. Therefore, in principle, we could use simple slab cell calculations to determine the fracture toughness for cracks on the polar plane, $\Gamma_{\mathrm{b}}^{\mathrm{p}}$. In practice, we use the absolute surface energies from Fig. 3.2, albeit enforcing the constraint of local, stoichiometry-preserving reconstructions.

For the nonpolar planes, the unreconstructed, relaxed surface was found to be the lowest-energy stoichiometry-preserving reconstruction for both GaN and AlN, consistent with previous calculations.[84,109,110] For the polar planes, we find that the lowest-energy reconstruction is a combination of the $2 \times 2$ cation vacancy on the $+c$ plane and a $2 \times 2$ cation adatom on the H3 site[84] of the $-c$ plane. This configuration is slightly lower in energy (by 0.3 J/m$^2$) than the N adatom/vacancy counterpart.

### 3.2.2 Fracture toughnesses of GaN and AlN

Our calculated values are presented in Table 3.1. We find that the trends in fracture toughness are the same for AlN and GaN, and the magnitudes are very similar. The fracture toughness of the $m$ plane is the lowest, though very similar to that of the $a$ plane. Our values for the nonpolar planes are larger by about 9 % than surface energies calculated previously[86,110] using traditional DFT with the local density approximation functional. The fracture toughness for the $c$ plane is substantially larger than that of the nonpolar planes.

### 3.2.3 Critical thickness for AlGaN on GaN

From the values of fracture toughness, we can determine a lower bound on the critical thickness [Eq. (3.7)] of AlGaN layers on GaN as a function of Al content. Elastic constants were taken from Ref. 5. We assume a Vegard's law dependence of fracture

**Table 3.1:** Fracture toughnesses in the brittle limit on the nonpolar and polar planes for GaN and AlN.

| | $\Gamma$ (J/m$^2$) | | |
|---|---|---|---|
| | $\{10\bar{1}0\}$ $m$ | $\{11\bar{2}0\}$ $a$ | $(0001)$+$(000\bar{1})$ $c$ |
| GaN | 3.88 | 3.97 | 6.38 |
| AlN | 4.81 | 4.97 | 6.31 |

toughness and elastic constants on Al content, though due to the similarity in the values for AlN and GaN, the linear interpolation gives a negligible change in the critical thickness compared to using only GaN or AlN values.

For layers grown on the $c$ plane, the tensile stress is equibiaxial and the elastic constants are isotropic in the $c$ plane so we use Eq. (3.7); since the $m$ and $a$ planes are so close in surface energy, we would predict cracking to occur on both of these two planes, which is consistent with observations of cracks on both nonpolar planes in GaN and AlGaN.[22,99–102] Figure 3.2.3 shows our calculated critical thicknesses of AlGaN versus Al fraction, compared with experimental observations from the literature. The two curves correspond to the different $Z$ values in Eq. (3.7) for surface and channel cracking. Most of these experimental studies measured the stress in the film with thickness (by *in situ* wafer curvature measurements),[100,103,108] and define the critical thickness as the onset of stress relaxation.

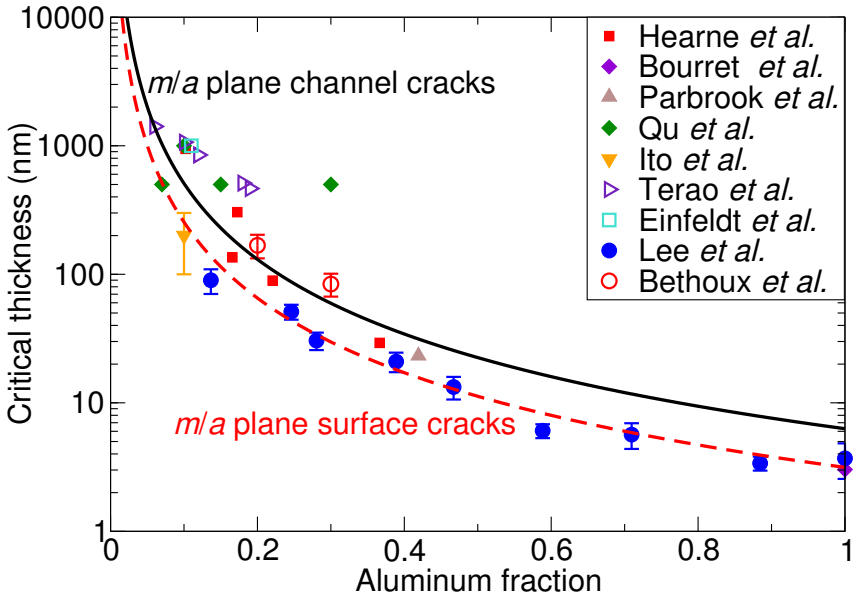


**Figure 3.5:** Calculated critical thickness for AlGaN on GaN versus Al content for *c* plane growth. Experimental values are included for comparison (Refs. Hearne [100], Bourret [106], Parbrook [107], Qu [105], Ito [104], Terao [108], Einfeldt [101], Lee [103], and Bethoux [102]).

Figure 3.2.3 shows that the onset of stress relaxation occurs at the thickness where our theoretical values predict surface cracks to form, indicating that stress relaxation occurs before the channeling process can begin, as was concluded by Lee *et al.*[103] Since it is unlikely that surface cracks could result in significant relaxation of strain energy,[99] this onset of relaxation before channeling is consistent with the proposed model of relaxation where cracks create local shear stresses and provide nucleation sites for MDs.[99,100,102,103] Some of the outlying studies on Fig. 3.2.3, such as Refs. 102 and 105, used the direct observation of cracks (using microscopy techniques) to indicate the critical thickness of the film,[99] which may cause a relative overestimation of the onset of stress relaxation since surface cracks may be difficult to observe before the channeling process begins.

For layers grown on a nonpolar plane, the film is no longer isotropic, and the stress is no longer equibiaxial, so Eq. (3.7) is not strictly valid. In order to determine $h_c$, we use the following simplifying assumptions: (1) The main mechanism of crack propagation is opening in the direction of stress, (2) the orthrotropic (anisotropic in orthogonal directions) film can be approximated as an isotropic plane with the elastic properties of the film perpendicular to the crack propagation direction, and (3) the nonequibiaixial strain can be approximated as plane strain where the film is fixed in the direction parallel to the crack propagation, strained perpendicular to the crack propagation and free perpendicular to the film.

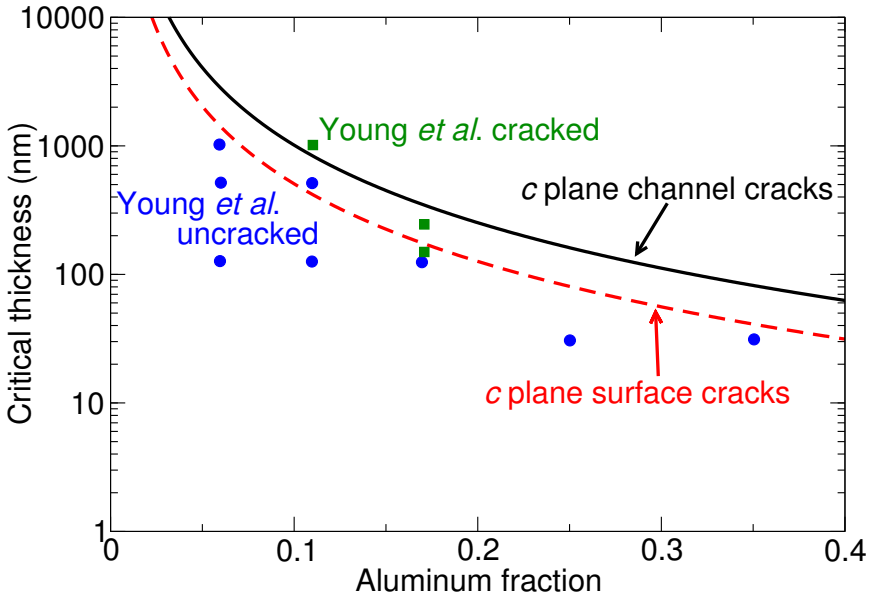


**Figure 3.6:** Calculated critical thickness for AlGaN on GaN versus Al content for nonpolar plane growth. Experimental values for $m$ plane growth from Young *et al.*[92] are included for comparison.

Due to the significantly larger amount of misfit strain in the $c$ direction, we find that the $c$ plane has the smallest critical thickness, in spite of the relatively large value of $\Gamma_{\mathrm{b}}^{\mathrm{p}}$ (Table 3.1). This is consistent with experimental observations of cracks propagating on the $c$ plane for nonpolar plane-grown AlGaN layers.[92] Figure 3.2.3 shows our predicted critical thickness as well as experimental observations from Young *et al.*[92] of whether or not the films had channel cracks. Our predicted critical thickness slightly overestimates the critical thickness for such cracks to form; this is because of our assumptions (plane strain and only crack opening parallel to the propagation direction) neglects the influence of shear stress contributions to crack propagation on the critical strain energy.

### 3.2.4 Conclusions

We have reported first-principles values for the brittle fracture toughness of GaN and AlN for cracks propagating on the $m$, $a$, and $c$ planes. Our results show that the fracture toughnesses are similar for GaN and AlN, and that the nonpolar planes have significantly smaller fracture toughnesses than the $c$ plane. Our brittle fracture toughnesses provide a lower bound for the critical thickness of AlGaN on GaN. For growth on the $c$ plane, our calculated fracture toughnesses indicate that cracks will form on the $m$ and $a$ planes, consistent with experimental observations, and support the model of MD-driven strain relaxation when surface cracks form. Also, our values indicate an overestimation of the critical thickness for cracking in many studies. For growth on the $m$ plane, we show that though the fracture toughness is higher, cracks are expected to propagate primarily on the $c$ plane, consistent with experimental observations.

# Chapter 4

# Spontaneous and piezoelectric polarization in III-nitrides

Because of the symmetry of the wurtzite (wz) crystal structure (space group $P6_3mc$), III-nitride materials exhibit spontaneous (sp) and piezoelectric (pz) dipole moments in the crystallographic $c$ direction,[4] which manifest themselves as electric fields in heterostructure layers and sheet charges at interfaces.

We have used first-principles calculations to determine the polarization properties of the III-nitrides. We will see in Section 4.2.1 that the pz constants can be measured experimentally, and calculated using a straightforward application of the theory in Section 4.1; therefore, much of this chapter will deal with the sp polarization constants. Direct experimental measurement of the magnitude of the sp polarization has not been possible

to date, and calculation of sp polarization is complicated by the necessity to choose a reference structure. The universal choice of reference structure for wz is zincblende (zb); we demonstrate in Section 4.2.2 that this choice does not allow consistent determination of the differences of sp polarizations between materials, which determines their physical manifestation.

In this Section 4.1 we present the derivation of the Modern Theory of Polarization and comment on several aspects of the theory. In Section 4.2 we report on our first-principles calculations for polarization constants in III-nitrides.

## 4.1 Modern Theory of Polarization

The ability of a material to be polarized, or to sustain a finite electric field, is one of the fundamental properties of an insulator. The phenomenological description of the polarization response of a dielectric material to an electric field is given by[112]

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}, \tag{4.1}$$

where $\chi_e$ is the susceptibility. Most textbooks still use the macroscopic, phenomenological definition of $\mathbf{P}$ as the dipole moment per unit volume of a material.[112]

This section outlines the microscopic theory of electronic polarization in crystals and demonstrates how it is calculated. The goal is to determine an expression for the intrinsic bulk polarization of a material. In other words, we would like to calculate the

average dipole moment for an infinite, periodic crystal. This may cause a reader with knowledge of electrostatics to cringe: the energy for an infinite system with a constant dipole density is not well defined. Even considering a single unit cell of a periodic crystal, the potential is only defined up to an arbitrary electric field [assuming the normal short circuit (constant $\mathbf{E}$) electrostatic boundary conditions];[113] to find the dipole density, we are looking for just such a constant electric field. These uncertainties have resulted in significant debate over whether a bulk definition of polarization is even possible for realistic materials. The Modern Theory of Polarization (MTP)[114,115] demonstrated that *differences* in $\mathbf{P}$ were indeed intrinsic bulk properties, and provided a simple procedure for calculating these $\Delta\mathbf{P}$'s within a first-principles electronic structure context.

We begin by considering a finite, neutral, charge density. The definition of dipole moment per volume provides the simple equation for the polarization

$$\mathbf{P} = \frac{1}{V}\int d\mathbf{r}\mathbf{r}\rho(\mathbf{r}), \tag{4.2}$$

where $\rho(\mathbf{r})$ is the charge density and $V$ is the volume. The value of $\mathbf{P}$ in Eq. (4.2) will have contributions from the interior of the charge distribution as well as from the specific morphology and properties of the surface of the charge distribution. For $\mathbf{P}$ to be a useful, intrinsic, material property, the contribution of the surface must be removed. However, disentangling the contributions of the bulk and surface to the total polarization is generally not possible. Therefore considering a finite charge distribution

[to which Eq. (4.2) can be applied] is not a good way to access the intrinsic properties of the material.

An obvious next step would be to consider, instead of a large, finite sample, an infinite bulk material where we define $\mathbf{P}$ as the dipole moment per volume of a unit cell

$$\mathbf{P} = \frac{1}{V_{\text{cell}}} \int_{\text{cell}} d\mathbf{r}\mathbf{r}\rho(\mathbf{r}). \tag{4.3}$$

However, it is easy to see that for a given periodic charge density, different choices of unit cell, Eq. (4.3) will give different values of the polarization; in fact, by choosing a sufficiently pathological cell, *any* value for $\mathbf{P}$ can be achieved.[116] Martin[117] demonstrated the inherent difficulty of this method. If we define a quantity called the microscopic polarization such that it satisfies the macroscopic empirical relation $\nabla \cdot \mathbf{P}_{micro}(\mathbf{r}) = -\rho(\mathbf{r})$ and integrate this over a cell volume we obtain two terms:

$$\mathbf{P} = \frac{1}{V_{\text{cell}}} \int_{\text{cell}} d\mathbf{r}\mathbf{r}\rho(\mathbf{r}) + \frac{1}{V_{\text{cell}}} \int_{\text{surface}} ds\mathbf{r}[\hat{n} \cdot \mathbf{P}(\mathbf{r})]. \tag{4.4}$$

The first is equivalent to Eq. (4.3), and the second depends on the surface of the cell. Therefore, knowledge of the charge density of the cell is not sufficient to define the polarization.

The MTP is motivated by a thought experiment for determining polarization not from charge density, but current. Consider a piezoelectric material; placing this material under strain will cause a redistribution of charge [Fig. 4.1(a)]. If contacts are placed on both sides of the material and connected, the deformation of the material will result in

a current flowing from one surface to the other [Fig. 4.1(b)]. The change in polarization with strain will be given by the integral of this polarization current as long as the deformation is carried out adiabatically, and the material remains insulating throughout (so that all of the current must flow through the wire and none can flow through the bulk of the material).[114] The MTP takes the situation in Fig. 4.1(b) as a starting point.

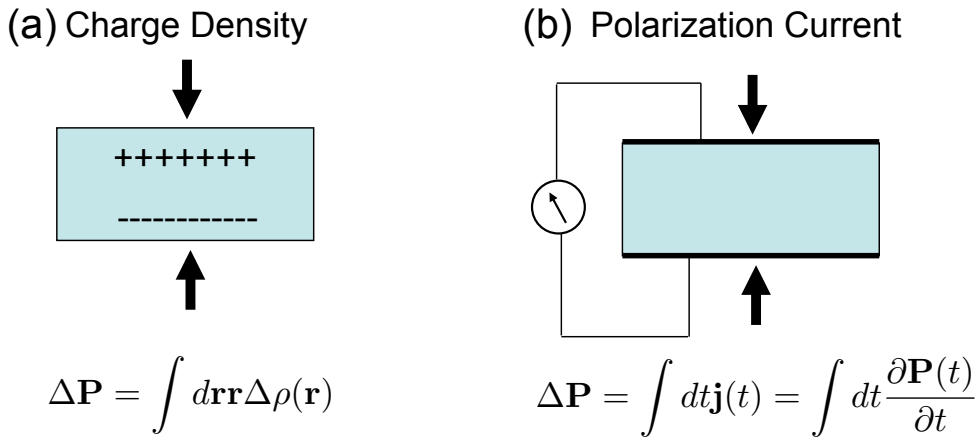


**Figure 4.1:** Schematic of piezoelectric measurement: a) charge density picture, and b) current density picture, which is the basis for the Modern Theory of Polarization.

We consider a unit cell of an infinite periodic crystal in a null electric field (as is the case in Fig. 4.1(b), as the field caused by the charges in (a) is shorted by the contacts), on which a general adiabatic, gap-preserving transformation (not necessarily a deformation as in Fig. 4.1) is applied. In the derivation below, we assume a KS-DFT context where all potentials (including pseudopotentials) are local, though the theory has been generalized for nonlocal pseudopotentials as well as XC functionals such as hybrids.[118]

Since we are considering an adiabatic transformation, we can parametrize the transformation with some generalized coordinate $\lambda$:

$$\Delta\mathbf{P} = \int_0^1 d\lambda \frac{\partial \mathbf{P}(\lambda)}{\partial \lambda}. \tag{4.5}$$

Because of the Born-Oppenheimer approximation, we can split this change in polarization into an ionic and electronic part:

$$\Delta\mathbf{P} = \Delta\mathbf{P}_{\text{ion}} + \Delta\mathbf{P}_{\text{el}} = \frac{e}{V}\sum_s Z_s \Delta\mathbf{R}_s + \int_0^1 d\lambda \frac{\partial \mathbf{P}_{\text{el}}(\lambda)}{\partial \lambda}, \tag{4.6}$$

where $s$ sums over the ions, $V$ is the volume of the cell, and $R$ are the nuclear coordinates. The ionic part of Eq. (4.6) is from simple electrostatics of point charges, so we will focus on the electric part. There are several paths to this derivation; I will follow a "historical" path of the development of the theory. We begin with work done by Resta.[114]

In DFT, the density is given by the modulus squared of the single-particle eigenfunctions of the Kohn Sham (KS) hamiltonian:

$$\rho_{el}(\mathbf{r}) = e\sum_i f_i |\phi_i(\mathbf{r})|^2, \tag{4.7}$$

so the electronic dipole moment of the eigenfunctions per unit volume is:

$$\mathbf{P}_{\text{el}}(\mathbf{r}) = \frac{e}{V}\sum_i f_i \langle \phi_i(\mathbf{r})|\mathbf{r}|\phi_i(\mathbf{r})\rangle, \tag{4.8}$$

where $f_i$ is the occupation of orbital $i$. Therefore the electronic polarization current can be written as:

$$\frac{\partial \mathbf{P}_{\text{el}}(\mathbf{r})}{\partial \lambda} = \frac{e}{V}\sum_i f_i \left( \langle \frac{\partial \phi_i(\mathbf{r})}{\partial \lambda}|\mathbf{r}|\phi_i(\mathbf{r})\rangle + \langle \phi_i(\mathbf{r})|\mathbf{r}|\frac{\partial \phi_i(\mathbf{r})}{\partial \lambda}\rangle \right), \tag{4.9}$$

where the dependence of the wave functions on $\lambda$ is suppressed (as will be done below). Since the KS eigenfunctions form a complete set at any $\lambda$, we can insert unity as such:

$$\frac{\partial \mathbf{P}_{\mathrm{el}}(\mathbf{r})}{\partial \lambda} = \frac{e}{V} \sum_i f_i \sum_j \left( \langle \frac{\partial \phi_i(\mathbf{r})}{\partial \lambda} | \phi_j(\mathbf{r}) \rangle \langle \phi_j(\mathbf{r}) | \mathbf{r} | \phi_i(\mathbf{r}) \rangle + c.c \right). \tag{4.10}$$

We can show with minor manipulations that for $j \neq i$:

$$\langle \frac{\partial \phi_i(\mathbf{r})}{\partial \lambda} | \phi_j(\mathbf{r}) \rangle = \frac{\langle \phi_i | \frac{\partial V_{\mathrm{KS}}}{\partial \lambda} | \phi_j \rangle}{\epsilon_i - \epsilon_j}. \tag{4.11}$$

To do this, we simply differentiate both sides of the single-particle KS equation with respect to $\lambda$:

$$\frac{d}{d\lambda} \langle \phi_j | H | \phi_i \rangle = \frac{d}{d\lambda} \langle \phi_j | \epsilon_i | \phi_i \rangle \tag{4.12}$$

$$\langle \phi_j | \frac{dH}{d\lambda} | \phi_i \rangle + \epsilon_j \langle \phi_j | \frac{d\phi_i}{d\lambda} \rangle + \epsilon_i \langle \frac{d\phi_j}{d\lambda} | \phi_i \rangle = \epsilon_i \left( \langle \frac{d\phi_j}{d\lambda} | \phi_i \rangle + \langle \phi_j | \frac{d\phi_i}{d\lambda} \rangle \right) \tag{4.13}$$

$$\langle \phi_j | \frac{d\phi_i}{d\lambda} \rangle = \frac{\langle \phi_j | \frac{dV_{\mathrm{KS}}}{d\lambda} | \phi_i \rangle}{\epsilon_i - \epsilon_j}, \tag{4.14}$$

where we used the fact that the kinetic energy in the KS Hamiltonian does not depend on $\lambda$, and that $H$ is hermitian. Assuming we do not have nonlocal terms in the Hamiltonian, we can write the dipole matrix element as:

$$\langle \phi_j | \mathbf{r} | \phi_i \rangle = -\frac{i\hbar}{m_e} \frac{\langle \phi_j | \mathbf{p} | \phi_i \rangle}{\epsilon_i - \epsilon_j}. \tag{4.15}$$

In the event we have nonlocal terms in the XC part of the Hamiltonian or in the pseudopotential, this commutation relation will have additional terms. These have been

worked out explicitly in Ref. 118. Using Eq. (4.15), we remove the dipole matrix element (which is ill defined in an infinite periodic system):

$$\frac{\partial \mathbf{P}_{\mathrm{el}}(\mathbf{r})}{\partial \lambda} = -\frac{ie\hbar}{m_e V}\sum_i f_i \sum_{j\neq i}\left(\frac{\langle\phi_i|\frac{dV_{\mathrm{KS}}}{d\lambda}|\phi_j\rangle\langle\phi_j|\mathbf{p}|\phi_i\rangle}{\left(\epsilon_i-\epsilon_j\right)^2}+c.c\right), \tag{4.16}$$

where we have explicitly removed the $j = i$ terms (which cancel with their complex conjugates). For a finite system, we have a discrete set of $\mathbf{k}$ points, and we can explicitly write Eq. (4.16) with the parametric dependence of the wave function on $\mathbf{k}$:

$$\frac{\partial \mathbf{P}_{\mathrm{el}}(\mathbf{r})}{\partial \lambda} = -\frac{ie\hbar}{m_e V}\sum_i f_i \sum_{\mathbf{k},\mathbf{k}'}\sum_{j\neq i}\left(\frac{\langle\phi_{i,\mathbf{k}}|\frac{dV_{\mathrm{KS}}}{d\lambda}|\phi_{j,\mathbf{k}'}\rangle\langle\phi_{j,\mathbf{k}'}|\mathbf{p}|\phi_{i,\mathbf{k}}\rangle}{\left(\epsilon_{i,\mathbf{k}}-\epsilon_{j,\mathbf{k}'}\right)^2}+c.c\right). \tag{4.17}$$

The momentum matrix element selects out only the terms $\mathbf{k} = \mathbf{k}'$, so we have:

$$\frac{\partial \mathbf{P}_{\mathrm{el}}(\mathbf{r})}{\partial \lambda} = -\frac{ie\hbar}{m_e V}\sum_i f_i \sum_{\mathbf{k}}\sum_{j\neq i}\left(\frac{\langle\phi_{i,\mathbf{k}}|\frac{dV_{\mathrm{KS}}}{d\lambda}|\phi_{j,\mathbf{k}}\rangle\langle\phi_{j,\mathbf{k}}|\mathbf{p}|\phi_{i,\mathbf{k}}\rangle}{\left(\epsilon_{i,\mathbf{k}}-\epsilon_{j,\mathbf{k}}\right)^2}+c.c\right). \tag{4.18}$$

Taking the thermodynamic limit makes the sum over $\mathbf{k}$ points into an integral over the Brillouin Zone (BZ):

$$\frac{\partial \mathbf{P}_{\mathrm{el}}(\mathbf{r})}{\partial \lambda} = -\frac{ie\hbar}{m_e 8\pi^3}\sum_i f_i \sum_{j\neq i}\int_{BZ} d\mathbf{k}\left(\frac{\langle\phi_{i,\mathbf{k}}|\frac{dV_{\mathrm{KS}}}{d\lambda}|\phi_{j,\mathbf{k}}\rangle\langle\phi_{j,\mathbf{k}}|\mathbf{p}|\phi_{i,\mathbf{k}}\rangle}{\left(\epsilon_{i,\mathbf{k}}-\epsilon_{j,\mathbf{k}}\right)^2}+c.c\right), \tag{4.19}$$

which is well defined in an infinite system. Also, Eq. 4.19 is *exact*, though it resembles a perturbation theory expression (and is referred to as such in Ref. 114). This expression, in principle, contains quantities that can be calculated: the eigenvalues, momentum matrix elements, and the derivative of the KS potential with $\lambda$. However Eq. (4.16) is unwieldy as it requires a sum over unoccupied states. King-Smith and Vanderbilt [115] provided a more intuitive form through some manipulations. We write the matrix elements

in terms of the cell periodic part of the of Bloch wave functions:

$$\begin{aligned}\langle\phi_{i,\mathbf{k}}|\mathbf{p}|\phi_{j,\mathbf{k}}\rangle &= \frac{m_e}{\hbar}\langle u_{i,\mathbf{k}}|\left[\frac{\partial}{\partial\mathbf{k}}, H_{\mathbf{k}}\right]|u_{j,\mathbf{k}}\rangle \\ \langle\phi_{i,\mathbf{k}}|\frac{\partial V_{\mathrm{KS}}}{\partial\lambda}|\phi_{j,\mathbf{k}}\rangle &= \langle u_{i,\mathbf{k}}|\left[\frac{\partial}{\partial\lambda}, H_{\mathbf{k}}\right]|u_{j,\mathbf{k}}\rangle,\end{aligned} \tag{4.20}$$

where $H_{\mathbf{k}}$ is the Bloch-Schrödinger Hamiltonian, as in Eq. (2.5), but with the KS potential:

$$H_{\mathbf{k}} = \frac{1}{2m_e}\left(-i\hbar\Delta + \hbar\mathbf{k}\right)^2 + V_{\mathrm{KS}}(\mathbf{r}). \tag{4.21}$$

The above relations can be shown through straightforward evaluation of the commutators. If we plug these into Eq. (4.19), and evaluate the commutators, we get:

$$\frac{\partial\mathbf{P}_{\mathrm{el}}(\mathbf{r})}{\partial\lambda} = -\frac{ie}{8\pi^3}\sum_i f_i \sum_{j\neq i}\int_{BZ} d\mathbf{k}\left(\langle\frac{\partial u_{i,\mathbf{k}}}{\partial\mathbf{k}}|u_{j,\mathbf{k}}\rangle\langle u_{j,\mathbf{k}}|\frac{\partial u_{i,\mathbf{k}}}{\partial\lambda}\rangle + c.c.\right). \tag{4.22}$$

Removing unity gives

$$\frac{\partial\mathbf{P}_{\mathrm{el}}(\mathbf{r})}{\partial\lambda} = -\frac{ie}{8\pi^3}\sum_i f_i \int_{BZ} d\mathbf{k}\left(\langle\frac{\partial u_{i,\mathbf{k}}}{\partial\mathbf{k}}|\frac{\partial u_{i,\mathbf{k}}}{\partial\lambda}\rangle - \langle\frac{\partial u_{i,\mathbf{k}}}{\partial\lambda}|\frac{\partial u_{i,\mathbf{k}}}{\partial\mathbf{k}}\rangle\right). \tag{4.23}$$

If we plug Eq. (4.23) into Eq. (4.6), we get

$$\Delta\mathbf{P}_{\mathrm{el}} = -\frac{ie}{8\pi^3}\sum_i f_i \int_{BZ} d\mathbf{k}\int_0^1 d\lambda\left(\langle\frac{\partial u_{\mathbf{k}i}}{\partial\mathbf{k}}|\frac{\partial u_{\mathbf{k}i}}{\partial\lambda}\rangle - \langle\frac{\partial u_{\mathbf{k}i}}{\partial\lambda}|\frac{\partial u_{\mathbf{k}i}}{\partial\mathbf{k}}\rangle\right). \tag{4.24}$$

We can integrate the second term by parts to get:

$$\begin{aligned}\Delta\mathbf{P}_{\mathrm{el}} = -\frac{ie}{8\pi^3}\sum_i f_i \int_{BZ} d\mathbf{k}\int_0^1 d\lambda\langle\frac{\partial u_{i,\mathbf{k}}}{\partial\mathbf{k}}|\frac{\partial u_{i,\mathbf{k}}}{\partial\lambda}\rangle - \\ \left\{\langle u_{i,\mathbf{k}}|\frac{\partial}{\partial k}|u_{i,\mathbf{k}}\rangle\right]_0^1 - \int_0^1 d\lambda\langle u_{i,\mathbf{k}}|\frac{\partial}{\partial\mathbf{k}}|u_{i,\mathbf{k}}\rangle\Bigg\},\end{aligned} \tag{4.25}$$

which reduces to:

$$\Delta\mathbf{P}_{\text{el}} = \frac{ie}{8\pi^3}\sum_i f_i \int_{BZ} d\mathbf{k}\left(\langle u_{i,\mathbf{k}}|\frac{\partial}{\partial\mathbf{k}}|u_{i,\mathbf{k}}\rangle\Big]_0^1 - \int_0^1 d\lambda \frac{\partial}{\partial\mathbf{k}}\langle u_{i,\mathbf{k}}|\frac{\partial}{\partial\lambda}|u_{i,\mathbf{k}}\rangle\right). \tag{4.26}$$

Since we have that $u_{i,\mathbf{k}}$ is periodic in $k$ space:

$$u_{i,\mathbf{k}} = e^{i\mathbf{G}\cdot\mathbf{r}}u_{i,\mathbf{k}+\mathbf{G}} \tag{4.27}$$

the term $\langle u_{i,\mathbf{k}}|\frac{\partial}{\partial\lambda}|u_{i,\mathbf{k}}\rangle$ is also periodic in $k$ space, and therefore the second term integrated over the BZ is zero. If we explicitly include the dependence of $u_{i,\mathbf{k}}$ on $\lambda$ as $u_{i,\mathbf{k}}^{(\lambda)}$, and assume constant occupation $f$ we have:

$$\Delta\mathbf{P}_{\text{el}} = \frac{ief}{8\pi^3}\sum_i^{occ}\int_{BZ} d\mathbf{k}\left(\langle u_{i,\mathbf{k}}^{(1)}|\frac{\partial}{\partial\mathbf{k}}|u_{i,\mathbf{k}}^{(1)}\rangle - \langle u_{i,\mathbf{k}}^{(0)}|\frac{\partial}{\partial\mathbf{k}}|u_{i,\mathbf{k}}^{(0)}\rangle\right). \tag{4.28}$$

We no longer have a sum over unoccupied states, and all that needs to be calculated is a $\mathbf{k}$ integral over the BZ. Including the ionic part, we have

$$\Delta\mathbf{P} = \frac{e}{V}\sum_s Z_s\Delta\mathbf{R}_s + \frac{ief}{8\pi^3}\sum_i^{occ}\int_{BZ} d\mathbf{k}\left(\langle u_{i,\mathbf{k}}^{(1)}|\frac{\partial}{\partial\mathbf{k}}|u_{i,\mathbf{k}}^{(1)}\rangle - \langle u_{i,\mathbf{k}}^{(0)}|\frac{\partial}{\partial\mathbf{k}}|u_{i,\mathbf{k}}^{(0)}\rangle\right). \tag{4.29}$$

We see immediately that we can split Eq. (4.29) into a difference between terms that come from a given $\lambda$ structure. These terms are the so-called "formal polarizations" of a structure:

$$\mathbf{P}_{formal}^{(\lambda)} = \frac{e}{V}\sum_s Z_s\mathbf{R}_s^{(\lambda)} + \frac{ief}{8\pi^3}\sum_i^{occ}\int_{BZ} d\mathbf{k}\langle u_{i,\mathbf{k}}^{(1)}|\frac{\partial}{\partial\mathbf{k}}|u_{i,\mathbf{k}}^{(1)}\rangle \tag{4.30}$$

An interesting observation about the electronic part of Eq. (4.30) is that the integral resembles a Berry phase, or the geometric phase that can result from integration over

a periodic manifold. A careful reader might point out that a Berry phase is defined in this way only if the manifold is periodic, and the BZ does not technically fall into this category; the Berry phase formalism has been generalized for just this case by Zak.[119]

An important comment about the above derivation results from the periodicity of the Bloch functions. We chose the form of the functions as in Eq. (4.27), but we could add any phase of the form $e^{i\beta(\mathbf{k})}$ such that

$$\beta(\mathbf{k}+\mathbf{G}) - \beta(\mathbf{k}) = 2\pi n, \tag{4.31}$$

where $\mathbf{G}$ is any reciprocal space lattice constant, or, equivalently

$$\beta(\mathbf{k}) = \beta_{\text{periodic}}(\mathbf{k}) + \mathbf{k}\cdot\mathbf{R}, \tag{4.32}$$

where $\beta_{\text{periodic}}$ is some function with the periodicity of the lattice and $\mathbf{R}$ is any real-space lattice vector. Consider if we had chosen a Bloch function of the form

$$\tilde{u}_{i,\mathbf{k}}(\mathbf{r}) = e^{i\beta}u_{i,\mathbf{k}}(\mathbf{r}). \tag{4.33}$$

If we plug this into the equation for the electronic part of the formal polarization we get

$$\tilde{\mathbf{P}}^{(\lambda)}_{\text{formal}} = \frac{ief}{8\pi^3}\sum_i^{\text{occ}}\int_{\text{BZ}} d\mathbf{k}\langle\tilde{u}^{(\lambda)}_{i\mathbf{k}}|\frac{\partial}{\partial\mathbf{k}}|\tilde{u}^{(\lambda)}_{i\mathbf{k}}\rangle = \mathbf{P}^{(\lambda)}_{\text{formal}} - f\sum_i\frac{eR}{V}, \tag{4.34}$$

where, $R$ is the modulus of any real-space lattice vector. Therefore, we see that the formal polarization from Eq. (4.30) is only well defined modulo a "quantum of polarization," $eR/V$. The implications of this fundamental uncertainty will be discussed further in Section 4.2.2.

Returning to Eq. (4.29), we see that the fundamental definition of polarization is defined as a difference. Even though we have *a posteriori* split up this difference in terms of formal polarizations, only the *difference* between formal polarizations is physically meaningful; every polarization needs to have a reference $\lambda = 0$ structure. Such a reference is straightforward in the case of piezoelectric polarization (the undeformed crystal, for example). However, for the case of spontaneous polarization (or pyroelectricity) the idea of a reference structure is not so straightforward. In order to address this, King-Smith and Vanderbilt[115] defined an "effective polarization" of a given structure as the difference in formal polarization of that structure and a high-symmetry reference (usually centrosymmetric). Again, this will be discussed further in Section 4.2.2.

The above derivation was conducted with the condition that the $\lambda = 0$ and $\lambda = 1$ structures were connected by an adiabatic, gap-preserving transformation. A useful generalization, called the interface theorem, was developed by Vanderbilt and King-Smith.[120] They showed that the difference in two materials' formal polarization would give the bound charge at an insulating interface between the two materials, as in the phenomenological relation that arises from the continuity of the displacement across an interface with no free charge:

$$\sigma_{\text{bound}} = (\mathbf{P}_{\text{f}}^{\lambda=1} - \mathbf{P}_{\text{f}}^{\lambda=0}) \cdot \hat{\mathbf{n}}. \tag{4.35}$$

This extends the physical applicability to, for example, interfaces between different materials, which cannot be connected by an adiabatic transformation but can have an insulating interface.

An interesting note about the above derivation is that instead of Bloch functions, it can be recast in terms of Wannier functions of the form

$$w_i(\mathbf{r}-\mathbf{T}) = |\mathbf{T}_i\rangle = \frac{V}{8\pi^3}\int_{\mathrm{BZ}} d\mathbf{k} e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{T})} u_{i,\mathbf{k}}. \tag{4.36}$$

The "Wannier center", which defines the center of mass of the Wannier function can be written as[120]

$$\langle \mathbf{T}_i|\mathbf{r}|0_j\rangle = i\frac{V}{8\pi^3}\int_{\mathrm{BZ}} d\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{T}}\langle u_{i,\mathbf{k}}|\frac{\partial}{\partial\mathbf{k}}|u_{i,\mathbf{k}}\rangle \tag{4.37}$$

Therefore, the electronic part of Eq (4.29) can be rewritten as

$$\Delta\mathbf{P}_{\mathrm{el}} = -\frac{ef}{V}\sum_{i}^{\mathrm{occ}}\left(\langle 0_i^{(\lambda=1)}|\mathbf{r}|0_i^{(\lambda=1)}\rangle - \langle 0_i^{(\lambda=0)}|\mathbf{r}|0_i^{(\lambda=0)}\rangle\right) \tag{4.38}$$

This interpretation demonstrates that the contribution of a band to the change in polarization can be simply expressed as a motion of that Wannier function associated with that band as if it was a classical charge.

A final comment is that the above derivation involves the wave functions of the material; the careful reader might find this at odds with the Hohenberg-Kohn theorem,[34] which states that all properties can be derived from the ground state density. In fact it can be shown[121] that the MTP is fully consistent with the Hohenberg-Kohn theorem,

though the above derivation provides a convenient implementation using wave functions explicitly.

## 4.2 Polarization constants in III-nitrides

For nitride films grown in the [0001] direction (i.e., the $+c$ direction), the polarization moment $P_3$ is given by the sum of the sp polarization, $P_{\rm sp}$, and the $z$ component of the pz polarization:[4]

$$P_3 = P_{\rm sp} + (\epsilon_1 + \epsilon_2)e_{31} + \epsilon_3 e_{33}, \tag{4.39}$$

where (in Voigt notation)[4] $\epsilon_i$ ($i$=1,2,3) is the strain in the $i$ direction and $e_{3i}$ are the corresponding piezoelectric constants. Experimentally, the determination of the separate sp and pz contributions to the total polarization moment in Eq. (4.39) is very difficult. Though, in principle, the pz polarization constants are fairly straightforward to explicitly measure or calculate,[122] there is a considerable spread in reported values.[123] As mentioned above, direct determination of sp polarization constants, is much more subtle, and has eluded explicit experimental determination. Therefore, theoretical values are required for interpretation of experiments or for performing Schrödinger-Possion simulations of heterostructures.

### 4.2.1 Piezoelectric constants

The calculation of piezoelectric constants is a straightforward application of the MTP using Eq. (4.29),[124] and several experimental and theoretical studies have been conducted for III-V nitrides.[123] There are three unique piezoelectric polarization constants in the wurtzite structure, though here we focus on $e_{33}$ and $e_{31}$, which are given by[125]

$$e_{33} = c_0 \frac{\partial P_3}{\partial c} + \frac{4ec_0}{\sqrt{3}a_0^2} Z^* \frac{du}{dc} \tag{4.40}$$

$$e_{31} = \frac{a_0}{2} \frac{\partial P_3}{\partial a} + \frac{2e}{\sqrt{3}a_0} Z^* \frac{du}{da} \tag{4.41}$$

where $Z^*$ is the axial component of the Born effective charge tensor given by

$$Z^* = \frac{\sqrt{3}a_0^2}{4e} \frac{\partial P_3}{\partial u} \tag{4.42}$$

We illustrate these elements schematically in Fig. 4.2.

Care must be taken concerning the electrostatic boundary conditions when considering the piezoelectric constants, as there are two piezoelectric tensors.[122] The "improper" piezoelectric tensor is given by

$$e_{ijk} = \frac{\partial P_i}{\partial \epsilon_j k}, \tag{4.43}$$

and is correct under open circuit ($\mathbf{D}$=0) boundary conditions. Equations (4.40) and (4.41) correspond to improper piezoelectric constants. The "proper" piezoelectric tensor is given by

$$\tilde{e}_{ijk} = \frac{\partial J_i}{\partial \epsilon_j k}, \tag{4.44}$$

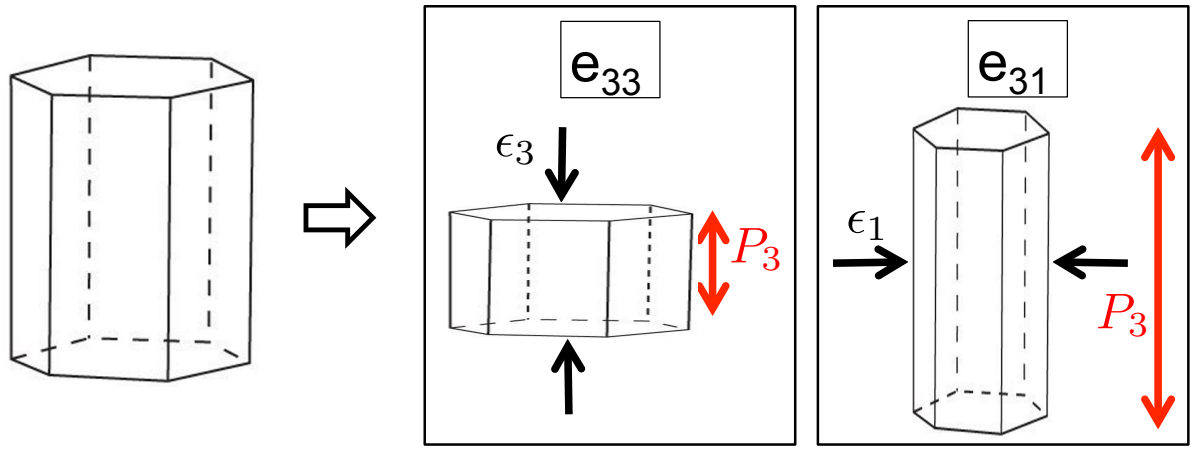


**Figure 4.2:** Schematic illustrating the definition of the piezoelectric constants in the wurtzite structure.

where $J_i$ is the polarization current. The constants are applicable to short circuit ($\mathbf{E} = 0$) electrostatic boundary conditions. The proper piezoelectric constants can be obtained from the improper ones as follows:[124]

$$\tilde{e}_{ijk} = e_{ijk} + \delta_j k P_i^{\mathrm{f}} + \delta_i k P_j^{\mathrm{f}} \tag{4.45}$$

where $P^{\mathrm{f}}$ is the formal polarization at zero strain. We report proper piezoelectric constants below.

There is considerable scatter in calculated values because these values depend on structural parameters of the crystal, which are different for different functionals. We perform calculations of the piezoelectric constants for the III-nitrides using the HSE hybrid functional,[46] which has been shown to give lattice parameters in very good agreement with experiment (Table 1.1). In addition, traditional DFT functionals such as those

based on the LDA and GGA predict InN to be a metal, whereas HSE correctly predicts the band gap.

**Table 4.1:** Proper piezoelectric polarization constants.

| | | DFT (GGA) | HSE | Prev. reported values[4] |
|---|---|---|---|---|
| GaN | $e_{31}$ | –0.409 ($C/m^2$) | –0.551 ($C/m^2$) | –0.22 to –0.55 ($C/m^2$) |
| | $e_{33}$ | 0.774 | 1.020 | 0.43 to 1.12 |
| AlN | $e_{31}$ | –0.589 | –0.676 | –0.38 to –0.74 |
| | $e_{33}$ | 1.473 | 1.569 | 1.29 to 1.94 |
| InN | $e_{31}$ | – | -0.604 | –0.38 to –0.74 |
| | $e_{33}$ | – | 1.238 | 1.29 to 1.94 |
| BN | $e_{31}$ | 0.381 | 0.282 | 0.27 |
| | $e_{33}$ | –1.046 | –0.952 | -0.85 |

The proper piezoelectric constants are given in Table 4.1. We see that AlN, GaN, and InN have similar piezoelectric behavior. BN however, has qualitatively different behavior, as the coefficients are opposite in sign to those of the other III-nitrides. This will be discussed further in Section 4.2.4.

### 4.2.2 Spontaneous polarization constants

For $P_{sp}$ in Eq. (4.39), the value of the sp polarization for the *equilibrium* structure of the bulk material is required, since strain effects are accounted for in the (proper) polarization constants in the second and third terms in Eq. (4.39).[124] For nitrides, the values that have been exclusively (and very widely) used for $P_{\rm sp}$ are those calculated by Bernardini *et al.*[125,126] We will show below that these values have been misused, and correct treatment of the contribution of sp polarization in these materials results in a significant difference when predicting bound sheet charge density at heterostructure interfaces. This difference can be as large as 40–100 % for a GaN/AlGaN interface (depending on Al content), and up to a factor of 5 for an InGaN/GaN interface. We provide revised values that are consistent with the current methods of implementation of polarization constants.

#### Effective spontaneous polarization

As mentioned above, only *differences* between formal polarizations of appropriate structures, $\lambda = 0$ and $\lambda = 1$, given by

$$\Delta\mathbf{P} = \mathbf{P}_{\rm f}^{(\lambda=1)} - \mathbf{P}_{\rm f}^{(\lambda=0)}, \tag{4.46}$$

are well defined, albeit modulo the "quantum of polarization", $eR/\Omega$, where $R$ is a lattice constant, $e$ is the electron charge, and $\Omega$ is the volume of the cell in which the formal polarization is measured.[116]

The choice of "appropriate" structures $\lambda = 0$ and $\lambda = 1$ rests on the two considerations mentioned in Section 4.1 to ensure that physical conclusions can be drawn from their formal polarization differences. First, if the two structures are connected by an adiabatic, gap-preserving deformation path,[114] then their difference in polarization [$\Delta\mathbf{P}$ in Eq. (4.46)] is given by Eq. (4.5), and corresponds to the zero-field, adiabatic displacement current. This quantity can in principle be determined experimentally and is the basis for the calculation of piezoelectric contents performed above. Second, from Section 4.1, Vanderbilt and King-Smith[120] showed that *if* an insulating interface can be constructed between two structures, then the difference in formal polarization gives the bound charge $\sigma_{\text{bound}}$ that builds up at the interface, as in the phenomenological relation from the continuity of the displacement field over an interface with no free charge [Eq. (4.35)]. Since there is no adiabatic path necessary between the two structures in this consideration, $\lambda = 0$ and $\lambda = 1$ can be different polymorphs of the same material (such as wz and zb structures of GaN) or different materials altogether (such as GaN and AlN); as long as they form an insulating interface, Eq. (4.35) will give the bound charge accumulation at the interface.

The above description highlights the fact that only *differences* in sp are well defined in the MTP. King-Smith and Vanderbilt[115] defined an "effective" spontaneous polarization of a structure $\lambda = 1$ as the $\Delta\mathbf{P}$ in Eq. (4.46) if the $\lambda = 0$ "reference" structure is a high-symmetry, centrosymmetric structure.[116] For the case of centrosymmetric crystals, however, the formal polarization need not vanish. This is because, as stated above, it is a multivalued vector field (as a result of the uncertainty modulo $eR/\Omega$), so it is possible for a nonzero value to be unchanged (modulo $eR/\Omega$) under the inversion operator. Nevertheless, high symmetry puts severe restrictions on the possible values of $\mathbf{P}_{\mathrm{f}}^{(\lambda=0)}$.[120]

While in principle effective polarization constants are still *differences* in formal polarization between $\lambda = 1$ and $\lambda = 0$ (reference) structures, in practice, they can be used to compare spontaneous polarization of different materials, if such materials share a reference structure with the same formal polarization. Such a comparison is then equivalent to the interface theorem of Ref. 120, and provides the $P_{\mathrm{sp}}$ in Eq. (4.39).

**Issues with zincblende as a reference structure for wurtzite**

Previous studies[125–127] have exclusively used zincblende (zb) (space group $F\bar{4}3m$) as a reference structure for calculating the spontaneous polarization. This structure is not centrosymmetric, though it has sufficient symmetry to preclude any spontaneous

polarization moment.[4] An insulating interface can be constructed between the wz and zb polytypes,[127] and therefore the zb structure is an appropriate reference structure.

However, there is an issue with using zb as a reference structure: it has a nonzero formal polarization $\mathbf{P}_{\mathrm{f}}^{\mathrm{zb}}$ (modulo $eR/\Omega$). Again, this is not inconsistent with the symmetry considerations because $\mathbf{P}_{\mathrm{f}}$ is a multivalued vector quantity, and can be nonzero while still remaining unchanged (modulo $eR/\Omega$) under the $F\bar{4}3m$ symmetry operations. Although $\mathbf{P}_{\mathrm{f}}^{\mathrm{zb}}$ can be calculated from an ideal point charge model,[120] and therefore does not depend on the species of the cation, the III-nitrides have different lattice constants so $\mathbf{P}_{\mathrm{f}}^{\mathrm{zb}}$ will be different for each material.

If we take the sp contribution to the interface polarization discontinuity, $\mathbf{P}_{\mathrm{int}}^{\mathrm{sp}}$, between different III-nitride materials ($a$ and $b$) as the difference in their effective polarizations (instead of their formal polarizations, as in the interface theorem), we have

$$\mathbf{P}_{\mathrm{int}}^{\mathrm{sp}} = \mathbf{P}_{\mathrm{eff}}^{a} - \mathbf{P}_{\mathrm{eff}}^{b} = \left[\left(\mathbf{P}_{\mathrm{f}}^{a,\mathrm{wz}} - \mathbf{P}_{\mathrm{f}}^{b,\mathrm{wz}}\right) - \left(\mathbf{P}_{\mathrm{f}}^{a,\mathrm{ref}} - \mathbf{P}_{\mathrm{f}}^{b,\mathrm{ref}}\right)\right], \tag{4.47}$$

where the first term on the right-hand side is the difference in formal polarizations of the wz polytype of material $a$ and $b$, and the second term is the difference between the formal polarizations of the reference structures. From the interface theorem, the polarization difference responsible for the physical effects (bound charges and fields) is just $\mathbf{P}_{\mathrm{f}}^{a,\mathrm{wz}} - \mathbf{P}_{\mathrm{f}}^{b,\mathrm{wz}}$, the first term in Eq. (4.47). Therefore the second term should be zero, as would be the case for equivalent reference structures; however for zb, as mentioned

above, the different formal polarizations of III-nitrides will result in this term being nonzero, representing a spurious contribution to $\mathbf{P}^{\mathrm{sp}}_{\mathrm{int}}$.

There is nothing intrinsically wrong with using zb as the reference structure, as long as one cancels the second term in Eq. (4.47), accounting for the difference in polarization between the zb reference structures. To our knowledge, however, this has not been properly implemented in the numerous previous evaluations of sp polarization, and it would require changes in the software for the many simulation tools that include modeling of polarization fields in heterostructures. In order to avoid such extensive changes, we advocate another approach, namely to choose a reference structure for which the formal polarization is explicitly zero (modulo $eR/\Omega$).

**Layered hexagonal structure as a consistent reference structure**

A straightforward choice for this reference structure is the layered hexagonal (h) structure (space group $P6_3/mmc$), as was used for hexagonal $P6_3mc$ $ABC$ materials.[128] This structure is centrosymmetric and is obtained by an adiabatic increase of the internal structural $u$ parameter from $u \approx 0.37 - 0.38$ of the wz structure to $u = 0.5$.

We perform first-principles calculations of $\mathbf{P}_{\mathrm{sp}}$ using density functional theory with the HSE hybrid functional[43,46] as implemented in the VASP code.[49] Hartree-Fock mixing parameters of 31 % for AlN and GaN, and 25 % for InN were used to correctly describe the band gaps and structural parameters of each material. Even though calculations

of formal polarization just involve integrating over the valence bands, which are well-described by traditional functionals, it is necessary to use HSE since LDA and GGA predict InN to be a metal (no band gap), precluding the calculation of the polarization constants if the gamma point is included in the $k$-point mesh (required for VASP). All calculations were performed on bulk primitive cells, with a $6 \times 6 \times 8$ Monkhorst-Pack[53] $k$-point mesh to sample the Brillouin zone, and an energy cut off of 600 eV for the plane-wave basis set to ensure convergence of the internal structural parameter $u$.[56]

We have calculated the electronic structure along this path to confirm that it is gap preserving (Fig. 4.3). Our calculations also show that the formal polarization of the h structure is zero (modulo $eR/\Omega$) for the III-nitrides (Fig. 4.3). We remind the reader that this was not guaranteed, since $\mathbf{P}_{\mathrm{f}}$ can be nonzero and still consistent with inversion symmetry, if the inversion operator changes $\mathbf{P}_{\mathrm{f}}$ by a multiple of $eR/\Omega$. We have therefore verified that the hexagonal phase is a reference structure for which there is no spurious term in Eq. (4.47).

The advantage of working with a reference structure with zero formal polarization is that we can focus on the formal polarizations of wz GaN, AlN, and InN; the only subtlety is that we must make sure that we are comparing values on the same "branch" with respect to the uncertainty modulo $eR/\Omega$. We can ensure this by correcting for any discontinuities (in the amount of a multiple of $eR/\Omega$) that may occur in our calculations of formal polarizations along the path between wz and h until we have smoothly varying

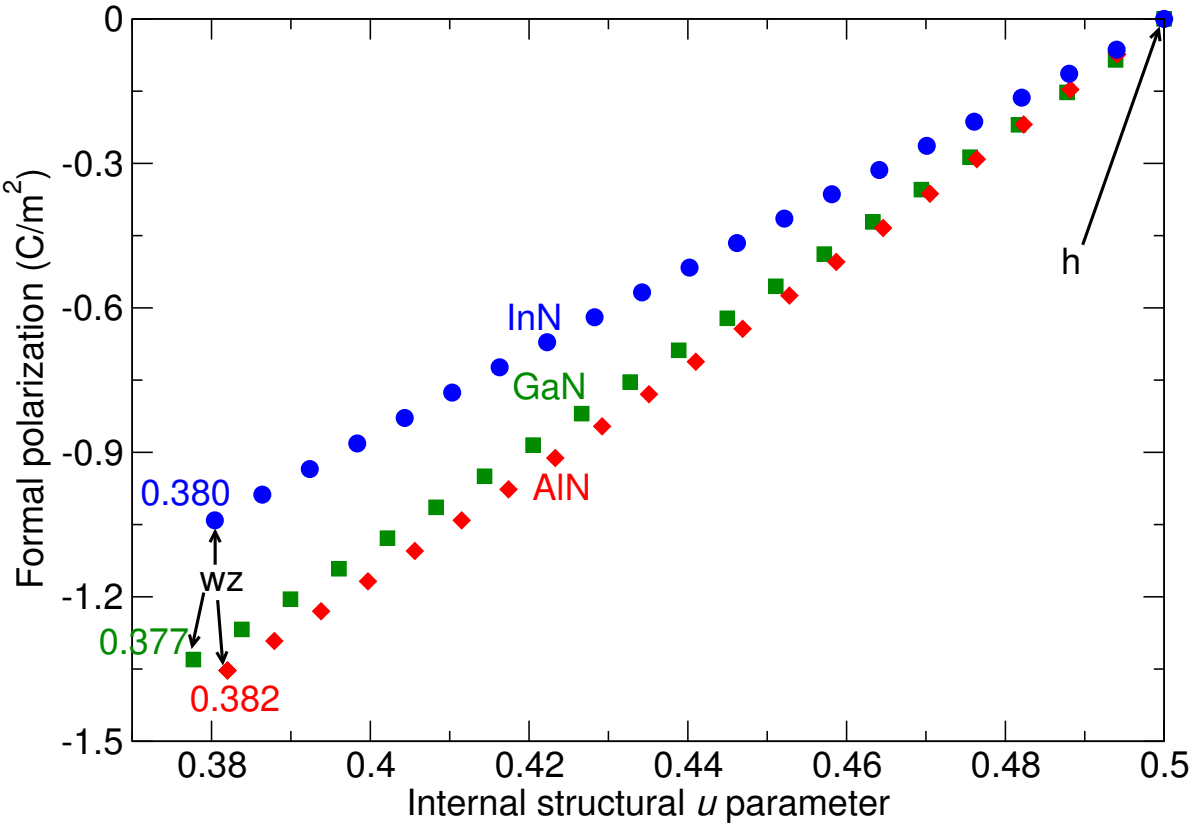


**Figure 4.3:** Formal polarization of InN, GaN, and AlN for structures with different internal structural $u$ parameters, interpolating between fully relaxed wz (indicated) and h ($u = 0.5$). All other lattice parameters were fixed at their relaxed wz values.

curves, as shown in Fig. 4.3. Note that the curves converge to a single value (equal to zero) at the h phase ($u = 0.5$); this would not be the case if zb were chosen as a reference.

The calculated spontaneous polarization coefficients for the wz structure using either h or zb as a reference are given in Table 4.2.2. The results obtained by Bernardini *et al.*[126] are listed for comparison; the GGA functional used in that work seems to provide

**Table 4.2:** Spontaneous polarization constants in units of $C/m^2$ of wurtzite (wz) GaN, AlN, and InN calculated using either the hexagonal (h, space group $P6_3/mmc$) or zincblende (zb, space group $F\bar{4}3m$) reference structures. Results from previous calculations[126] that used the zb reference are listed for comparison.

| material | h | zb | zb, prev.[1] |
|---|---|---|---|
| GaN | -1.330 | -0.035 | -0.034 |
| AlN | -1.353 | -0.090 | -0.090 |
| InN | -1.041 | -0.053 | -0.042 |

results that are very close to those obtained with HSE. Table 4.2.2 also shows, however, that the choice of reference structure makes a big difference. While the sign of all the spontaneous polarization constants remains the same, their magnitude is much larger in case of the h reference; recall that these numbers represent the bare formal polarization of the wz materials. We observe that it is not just the absolute values, but also the relative differences between the calculated polarization constants of the three materials that differ from the previously published values.[126]

### 4.2.3 Influence of revised spontaneous polarization constants on experimental observations

Because of the important impact of polarization on device performance and design, a plethora of experimental studies have been aimed at determining the effects of po-

larization at GaN/AlGaN and GaN/InGaN heterostructures. In Figs. 4.2.3 and 4.2.3, we compare the polarization sheet charge predicted based on linear interpolation of our spontaneous polarization constants as well as those calculated by Bernardini *et al.*,[126] with reported experimental values. Since experiment observes the *total* polarization [Eq. (4.39)], the comparison requires knowledge of the piezoelectric polarization; the shaded areas represent the range of the reported experimental and theoretical pz polarization constants.[123]

For the AlGaN/GaN system, there are two basic strategies for determining polarization effects. The first is to measure the 2DEG density at the AlGaN/GaN interface in a HEMT-like structure (GaN channel, AlGaN barrier); from this, the bound interface charge $\sigma_{\text{bound}}$ can be derived.[129] The 2DEG density can be determined either by Hall effect[130–145] or capacitance-voltage[129,146–148] measurements. We see from Fig. 4.2.3 that all of the largest sheet charge densities were derived from Hall measurements; we suggest this might be due to the fact that Hall measurements can be affected by parallel conduction pathways in underlying layers. Since background $n$-type doping in bulk GaN and AlGaN can be as high as $\sim 10^{18}\text{cm}^{-3}$, this could result in a spurious overestimate of $n_s$.[136,149–151]

The other strategy for determining the polarization effects is to directly measure the polarization *field* in an AlGaN/GaN/AlGaN quantum well (QW) structure. This field can be probed by varying QW width[151–157,157–166] or external biases[167–169,169,170]

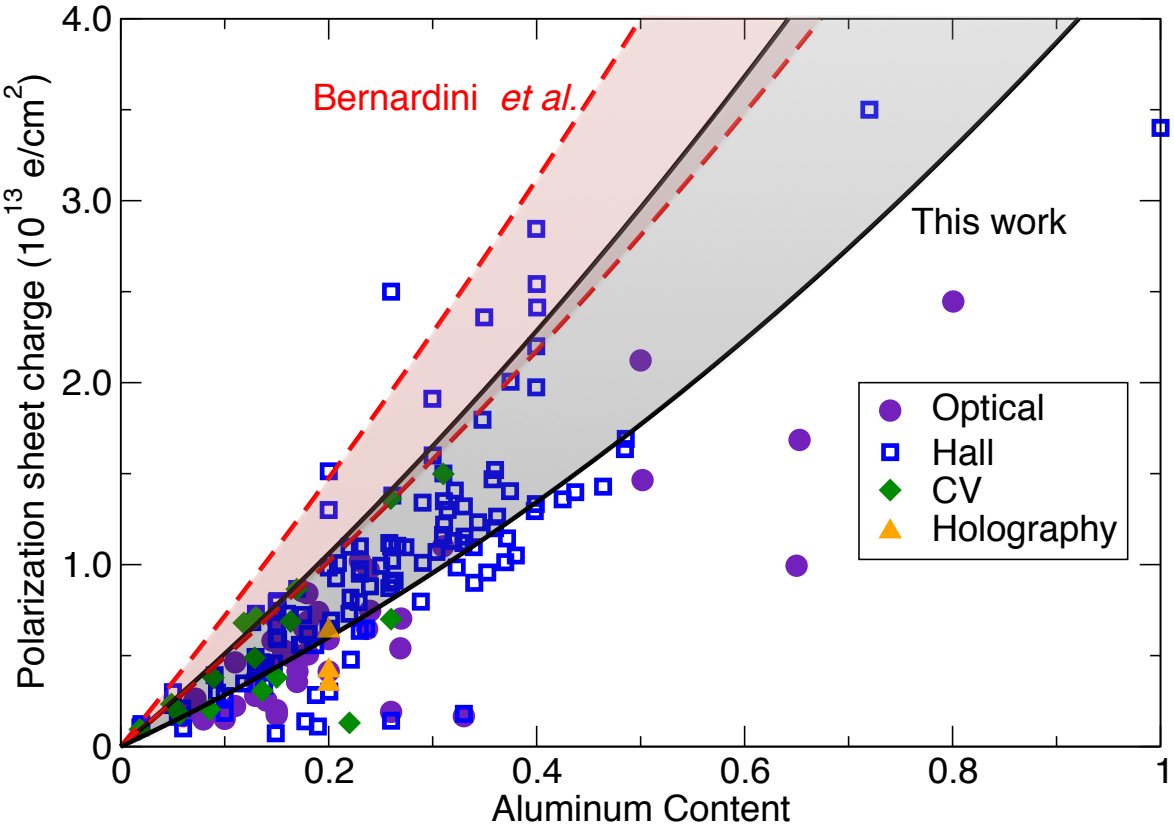


**Figure 4.4:** Values for polarization sheet charges at the AlGaN/GaN interface as a function of Al content predicted from the spontaneous polarization constants calculated in this work and from those reported in Ref. 126, compared with experimental values from the literature. The shaded regions indicate uncertainty in published piezoelectric constants.

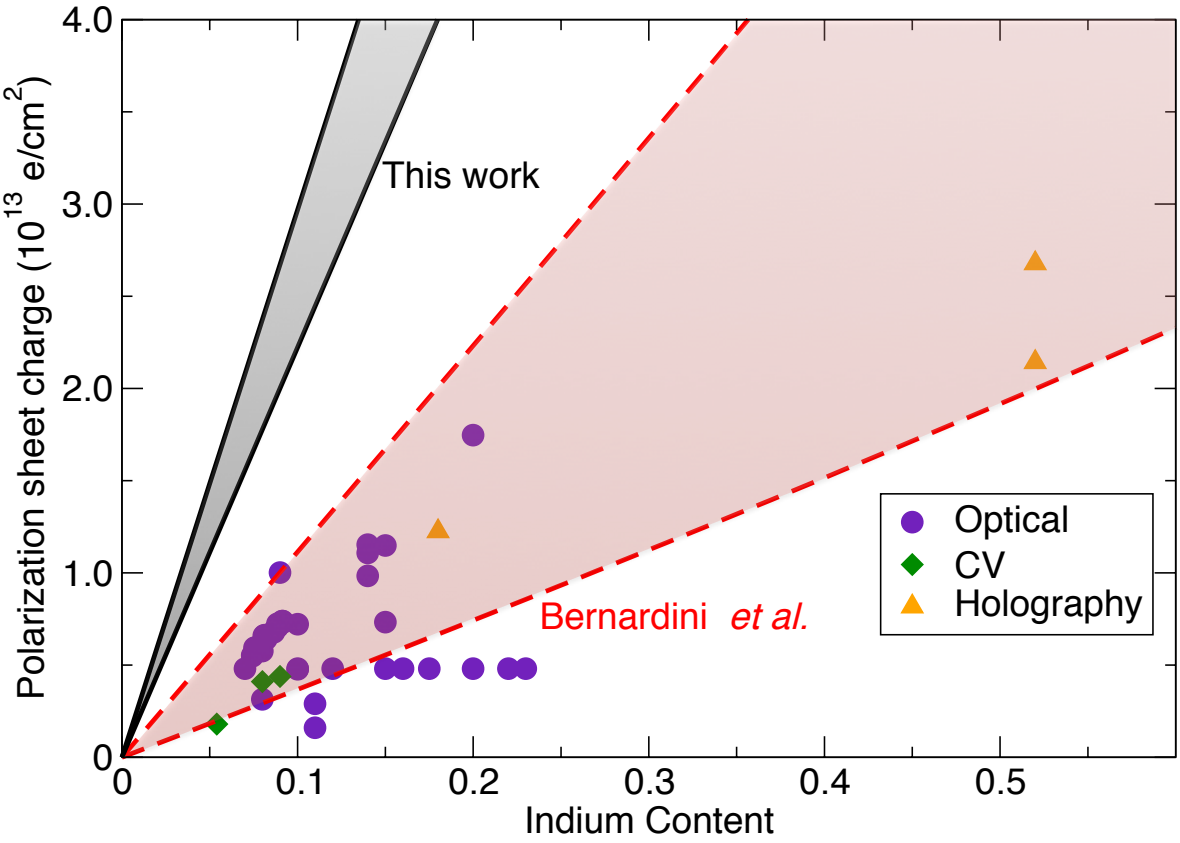


**Figure 4.5:** Values for polarization sheet charges at the InGaN/GaN interface as a function of In content predicted from the spontaneous polarization constants calculated in this work and from those reported in Ref. 126, compared with experimental values from the literature. The shaded regions indicate uncertainty in published piezoelectric constants.

and measuring the change in the optical properties of the QW. Other studies have been based on electron holography,[171] where cross-sectional transmission electron microscopy is conducted on AlGaN/GaN heterostructures to provide spatial resolution (in the growth direction) of the electric field in the sample. For the purposes of Fig. 4.2.3 we have converted these fields to bound sheet charge densities (lumped together under "optical") using a relative dielectric constants for GaN of 10;[172] the results turn out to be in good agreement with our calculated polarization constants. The significant scatter in the data is probably a result of the fact that these measurements usually rely on Schrödinger-Poisson solvers to determine the field magnitude from the measured optical properties. Uncertainties in input parameters to these models such as well widths, compositions and composition profiles,[173,174] incomplete strain relaxation in buffer layers, partial relaxation in the barrier layer,[175] and background doping could result in quantitative differences.

For the InGaN/GaN system, most experimental studies have applied the optical techniques outlined above to directly measure polarization fields in GaN/InGaN/GaN QWs (Fig. 4.2.3).[148,167,176–188] In addition, there have also been studies using time-resolved PL to measure shifts due to screening of the polarization field by photoexcited or electrically injected carriers.[189,190] We see from Fig. 4.2.3 that our calculated bound charge densities from these fields (assuming a relative dielectric constant for InN of 15[191] and a linear interpolation for the dielectric constant of InGaN) is significantly larger than the

measured value. Superficially, the experimental values seem to be in better agreement with the predictions based on the values of Bernardini *et al.*,[125,126] which probably contributed to the general acceptance of these values. However, we have shown above that our present values (using the h reference structure) are correct on theoretical grounds, and should be used in device simulations; we argue that the interpretation of experimental observations of polarization in InGaN/GaN systems is responsible for this apparent discrepancy.

Again, Schrödinger-Poisson modeling is typically used to interpret the measured optical properties, and the same uncertainties arise as discussed in the case of AlGaN/GaN. In fact, deviations from an ideal QW structure are expected to be much more significant for InGaN/GaN because of the large lattice mismatch and the large difference in optimal growth temperatures for GaN and InGaN. Wu *et al.*[192] recently reported atom-probe data showing compositional fluctuations and diffuse boundaries in a commercial blue LED; they demonstrated that the average In profile in the growth direction more closely resembled a gaussian than a square well, with a peak In content lower than the nominal value by 2%. They also showed that taking the well profile and In fluctuations into account, the predicted turn-on voltage was similar to that of a square well with significantly lower (less than half of the accepted Bernardini *et al.*[125,126] values) polarization constants, and in better agreement with experimental curves.

Such a significant deviation from ideal wells will also complicate the interpretation of changes in optical properties with well width or bias voltage used to deduce polarization constants. Firstly, the In composition of the QW is often deduced from the wavelength of absorption/luminescence properties; since different In profiles will result in different amounts of quantum confinement, the assumption of an unrealistic, ideal well profile will cause errors in the inferred In (as found in Ref. 192). Therefore, the polarization constants deduced will not correspond to the InGaN alloy for which they are reported. Also, a rounded well profile will reduce the dependence of the transition energies on field and well thickness compared to a square well.[193] Therefore, the assumption of a square well when interpreting such dependencies would result in an *underestimation* of the polarization discontinuity in the heterostructure.

In addition, calculations by Bernardini and Fiorentini[194] have suggested that there is a significant deviation from Vegard's law when interpolating between GaN and InN polarization constants, much more so than for AlGaN alloys. This bowing does not depend on reference structure for the spontaneous polarization, and therefore would also cause a significant reduction in the interpolated sp polarization constants in this work.

### 4.2.4 Polarization of wurtzite BN

We also calculate the sp polarization of BN, and summarize the results, along with the pz constants mentioned above, in Table 4.3. We also include previous LDA calculations reported in Ref. 195. These values have been published in Ref. 196.

**Table 4.3:** Calculated polarization constants for wurtzite BN. All values are in $C/m^2$.

| | This work | Previous calculations[195] |
|---|---|---|
| $e_{33}$ | –1.068 | –0.85 |
| $e_{31}$ | 0.282 | 0.27 |
| $P_{sp}$ (zb reference) | –0.012 | |
| $P_{sp}$ (h reference) | –2.174 | |

We can use these calculations to explore the alloying of BN with AlGaN, which has been proposed to modify the polarization properties of barriers. The signs of the corresponding pz coefficients between BN and AlGaN are opposite in sign, so, in principle, the polarization of BAlN or BGaN alloys will be less sensitive to strain. However the magnitude of the sp polarization moment of BN (using the h reference) is much larger than that of AlN. This would result in a significantly larger bound polarization charge at, for example, a GaN/BAlGaN interface.

Figure 4.6 shows an estimate for this increase by interpolating the polarizations between AlN, GaN, and BN using Vegards law. We assume BAlGaN is coherently

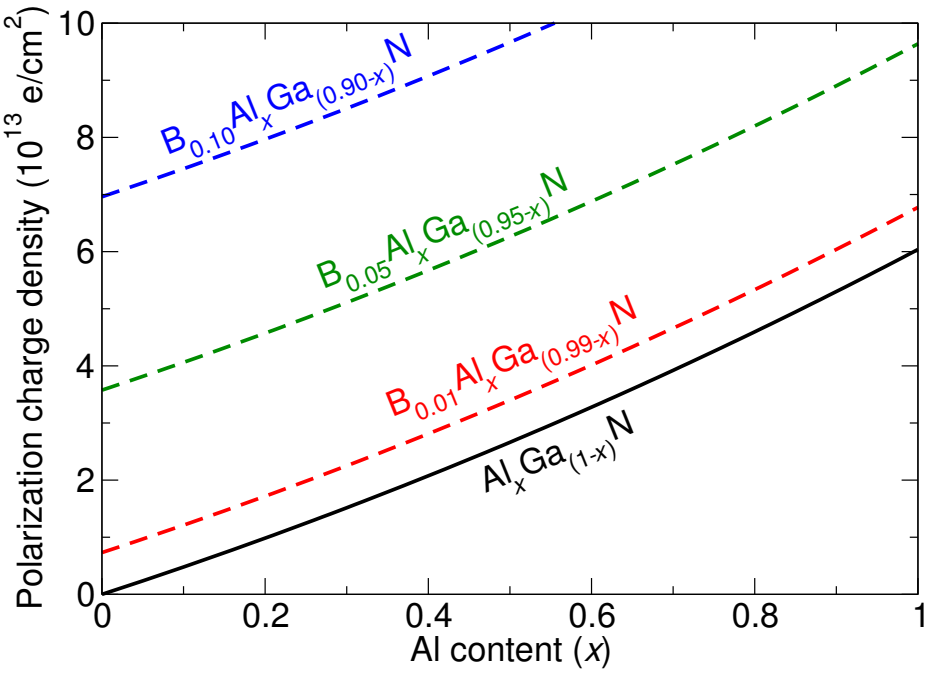


**Figure 4.6:** Increase in polarization bound charge at BAlGaN/GaN interface by alloying with different amounts of BN.

strained on GaN, and elastic constants are interpolated between AlN,[5] GaN,[5] and BN.[197] Because of the significant bowing expected for BAlGaN alloy properties,[16] the curves in Fig. 4.6 represent an upper bound on the polarization charge enhancement. We see that only a few percent BN in the alloy is sufficient to influence the polarization properties of the alloy and strongly enhance the polarization discontinuity at the interface with GaN. We note that the magnitude of the sp polarization with respect to the h reference is significantly larger than if zb is used as a reference.

### 4.2.5 General trends in spontaneous polarization in III-nitrides

We can observe an interesting trend when we add the formal polarization of wz-BN to Fig. 4.3 (see Fig. 4.7), namely that the magnitude of the formal polarization increases for decreasing volume. This suggests that the origin of the sp polarization (loosely speaking) is structural and not "chemical." In fact, if we calculate the polarization of all the III-nitrides with the lattice constants of GaN, we see that they fall on the same line (Fig. 4.8). We can understand this *a postiori* by realizing that the valence bands of all of the compounds, which enter into the expression for the electronic part of the sp polarization Eq. (4.29) are similar, as they are derived from N $p$-states.

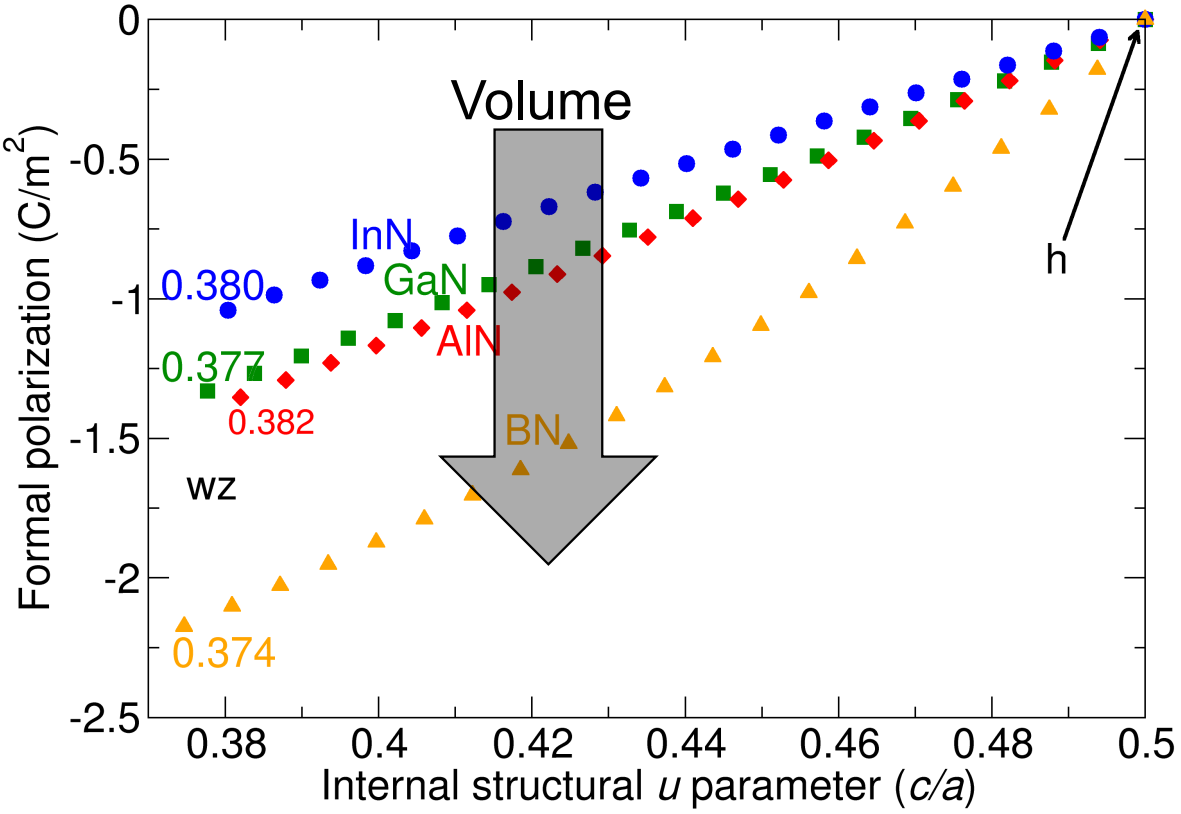


**Figure 4.7:** Formal polarization of InN, GaN, AlN, and BN for structures with different internal structural $u$ parameters, interpolating between fully relaxed wz (indicated) and h ($u = 0.5$). All other lattice parameters were fixed at their relaxed wz values.

## 4.3 Conclusions

We have calculated piezoelectric constants and revisited the values for the spontaneous polarization constants in the III-nitrides. We have demonstrated that the use of the zb reference structure results in spurious contributions to the polarization discontinuities at III-nitride heterostructure interfaces, and provided values with a consistent (h) reference structure to be used in simulations and to interpret experimental observations.

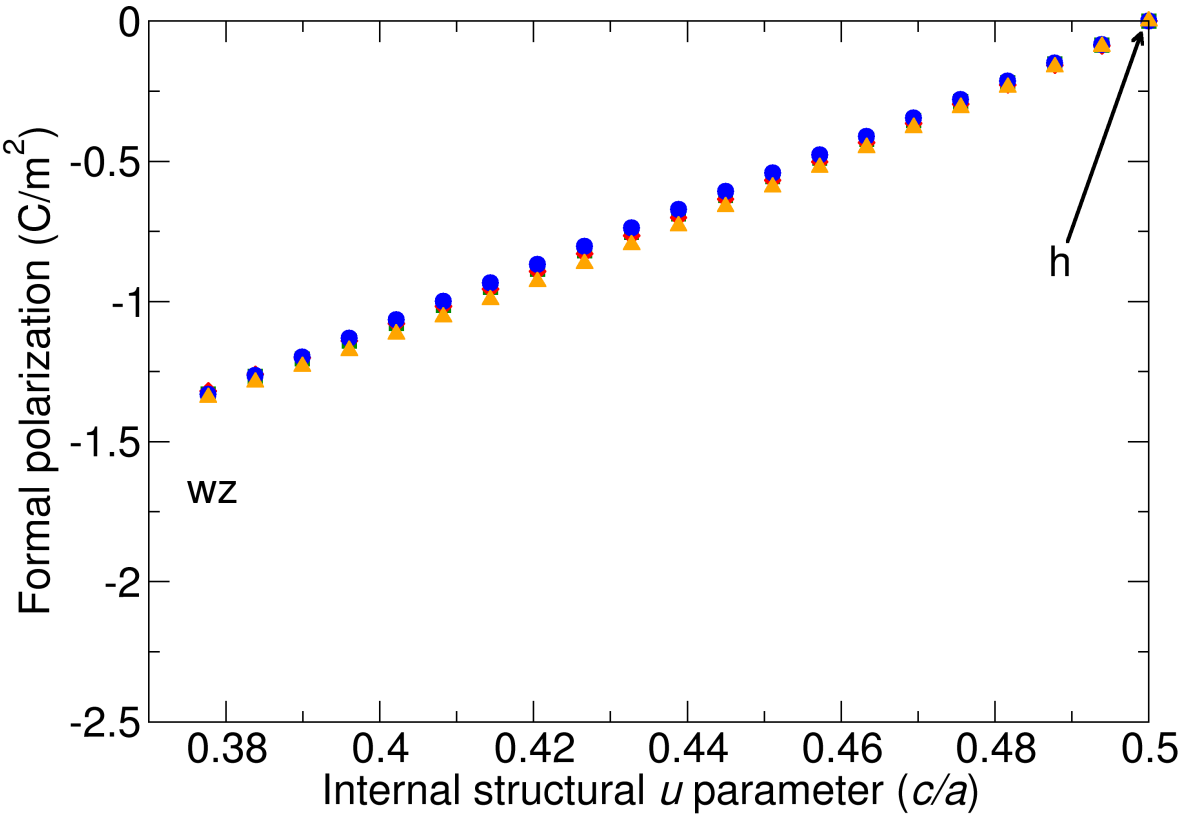


**Figure 4.8:** Formal polarization of InN, GaN, AlN, and BN for structures with different internal structural $u$ parameters, interpolating between fully relaxed wz (indicated) and h ($u = 0.5$). All other lattice parameters were fixed at the GaN values.

Our values are within the scatter of experimental observations for the AlGaN/GaN system, but result in larger polarization fields when compared to experiment for the InGaN/GaN system. This is most likely due to the nonideality of InGaN QW structures. We have attributed the differences in sp polarization between the nitrides to differences in volume. Finally, we have determined the influence of BN alloying with AlGaN for barrier applications.

# Chapter 5

# Summary and future directions

In this work, we employed first-principles techniques to investigate properties of III-nitride materials relevant for growth and device fabrication. We revisited several properties that were assumed to be well known and provided new insights.

We explored the influence of strain on the electronic structure of GaN and AlN, and found that the trends of the electron effective mass with strain differed from the expected behavior, as determined by $\mathbf{k} \cdot \mathbf{p}$ semi-emperical modeling. We demonstrated that this was due to the incorrect assumption that the momentum matrix elements between bands remains constant with strain. This work highlights the power of first-principles density functional theory calculations for testing and parametrizing semi-empirical models that may be then applied in situations where a full first-principles treatment is not computationally possible. We showed that first-principles calculations

of the momentum matrix elements needed for $\mathbf{k} \cdot \mathbf{p}$ theory could be very fruitful and could help in an expansion of the capabilities of $\mathbf{k} \cdot \mathbf{p}$ theory.

In addition, with the motivation of understanding growth morphologies and cracking, we have studied the energetics of surfaces in GaN. Using a model based on the zincblende polytype of GaN, we have overcome the symmetry (or lack thereof) constraints on defining absolute surface energies for the polar (0001) and $(000\bar{1})$ *c* planes, which are the most common growth planes for nitrides. We have also calculated surface energies of the nonpolar $\{10\bar{1}0\}$ *m* and $\{11\bar{2}0\}$ *a*. Taking into account both native reconstructions and those involving hydrogen, we have determined the stable surfaces for a range of experimentally relevant growth conditions.

Using these values of the surface energies and a simple model of cracking, we have determined toughnesses for brittle fracture on the polar and nonpolar planes. Combining this with surface energies for AlN, we calculate the critical thickness for cracks to form of AlGaN grown on GaN. We find that our results are in good agreement with available experimental data for growth on the *c* and *m* planes, and will be valuable when determining thickness limitations for more complicated situations such as semipolar growth.

We will explore the connection between our calculations of surface energy and experimental results for additional conditions to those presented in this study; an example is ammonothermal growth conditions for the growth of bulk GaN. Also, our method to

determine separate polar surface energies can be applied to other III-nitrides, as well as other wurtzite materials such as ZnO and hexagonal polytypes of SiC.

Finally, we have revisited the polarization properties of the III-nitrides. We have shown that the spontaneous polarization constants that have been traditionally used, which are based on the zincblende reference structure, have been misinterpreted. This is due to the fact that the zincblende structure of the nitrides has a nonzero *formal* polarization, which enters as a spurious term when comparing the effective spontaneous polarizations of different materials. We show that the hexagonal, layered structure, which we find to have zero formal polarization, provides a consistent reference structure for the III-nitrides. We demonstrated that the revision of the spontaneous polarization constants has a significant effect on the parameters for device modeling as well as interpretation of experimental data. Also, we have calculated the polarization properties of BN and demonstrated how it can be alloyed with GaN, AlN, or AlGaN to tune the polarization properties of the alloy.

These new values for the polarization constants will be disruptive in the context of device modeling of III-nitride devices, requiring a reevaluation of interpretations from such models. Our calculations of polarization coefficients also emphasize the need to take into account the effects of nonidealities in device structures, and we are currently studying how nonabrupt interfaces influence the analysis of previous experimental observations of polarization in III-nitrides.

# Appendices

# Appendix A

# Reading List

Below is a list of references for various aspects of first principles calculations. I attempted to provide several references for each topic; the book chapters will be more useful for beginners, but I also included the original papers, which are good to read once familiar with the basic principles. The underlined topics are (generally) in order of what should be reviewed. Within each topic, the references are also (generally) in order of least to most advanced.

## A.1 Books

***ABCs of DFT***

**K. Burke**

`http://dft.uci.edu/sites/default/files/g1.pdf`

***Density Functional Theory: A Practical Introduction***

**D. Sholl and J. A. Steckel**

John Wiley & Sons, 2009.

***Electronic Structure: Basic Theory and Practical Methods***

**R. M. Martin**

Cambridge University Press, 2004.

***A Primer in Density Functional Theory***

**C. Fiolhais, F. Nogueira, and M. Marques**

Springer, 2003

## A.2 Additional references

***Theory and Practice of Density-Functional Theory***

**P. Blöchl**

`http://arxiv.org/pdf/1108.1104v1.pdf`

***Density Functional Theory: An Advanced Course***

**E. Engel and R.M. Dreizler**

Springer Berlin Heidelberg, 2011

http://www.springerlink.com/content/978-3-642-14089-1/contents/

***Nobel Lecture: Electronic structure of matter–***

***Wave functions and density functionals***

**W. Kohn**

http://rmp.aps.org/abstract/RMP/v71/i5/p1253_1

***The density functional formalism, its applications and prospects***

**R. O. Jones and O. Gunnerson**

http://rmp.aps.org/abstract/RMP/v61/i3/p689_1

***A bird's-eye view of density-functional theory***

**K. Capelle**

http://arxiv.org/abs/cond-mat/0211443

## A.3 Specific topics

Discussion of specific topics in the books listed above.

### A.3.1 Introduction/Motivation

**K. Burke**

Section I: Background

**D. Sholl and J. A. Steckel**

Chapter 1: What Is Density Functional Theory?

**R. M. Martin**

Chapter 1: Introduction

Chapter 2: Overview

Chapter 3: Theoretical Background

Chapter 4: Periodic Solids and Electron Bands

### A.3.2 Density Functional Theory

**K. Burke**

Section II: Basics

**R. M. Martin**

Chapter 6: Density functional theory: foundations

Chapter 7: The Kohn-Sham auxiliary system

**C. Fiolhais, F. Nogueira, and M. Marques**

Chapter 1: Density Functionals for Non-relativistic Coulomb Systems

**Journal articles**

P. Hohenberg, W. Kohn, "Inhomogeneous Electron Gas," Phys. Rev **136**, 864 (1964).

W. Kohn and L.J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects" Phys. Rev. **140**, 1133 (1965).

### A.3.3 XC functionals

**K. Burke**

Chapter 8: Local Density Approximation

Chapter 13: Adiabatic Connection (basis for hybrids)

Section IV: Beyond LDA

**R. M. Martin**

Chapter 8: Functionals for exchange and correlation

**C. Fiolhais, F. Nogueira, and M. Marques**

Section 1.6: Local, Semi-local and Non-local Approximations

**Journal articles**

J.P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," Phys. Rev. Lett **77**, 3865 (1996).

A.D. Becke, "A new mixing of Hartree-Fock and local density–functional theories," J. Chem. Phys. **98**, 1372 (1993).

K. Burke, M. Ernzerhof, and J.P. Perdew, "The adiabatic connection method: a non-empirical hybrid," Chem. Phys. Lett. **265**, 115 (1997).

J. Heyd, G. Scuseria, and M. Ernzerhof, "Hybrid functionals based on a screened coulomb potential," J. Chem. Phys. **118**, 8207 (2003).

J. P. Perdew, M. Ernzerhof, and K. Burke,"Rationale for mixing exact exchange with density functional approximations, J. Chem. Phys. **105**, 9982 (1996). D. M. Ceperley and B. J. Alder, "Ground State of the Electron Gas by a Stochastic Method," Phys. Rev. Lett. **45**, 566 (1980).

J. P. Perdew and A. Zunger, "Self-interaction correction to density-functional approximations for many-electron systems," Phys. Rev. B, **23**, 5048 (1981).

### A.3.4 Pseudopotentials

**R. M. Martin**

Chapter 11: Pseudopotentials

**C. Fiolhais, F. Nogueira, and M. Marques**

Section 6.3: Pseudo-potentials

**Journal articles**

D. Vanderbilt, "Soft self-consistent pseudopotentials in a generalized eigenvalue formalism" Phys. Rev. B **41**, 7892 (1990).

P. Blöchl, Projector augmented-wave method, Phys. Rev. B **50**, 953 (1994).

### A.3.5 Practical aspects of running calculations

**D. Sholl and J. A. Steckel**

Chapter 2: DFT Calculations for Simple Solids

Chapter 3: Nuts and Bolts of DFT Calculations

Chapter 10: Accuracy and Methods beyond Standard Calculations

**R. M. Martin**

Chapter 12: Plane Waves and grids: basics

Chapter 13: Plane waves and grids: Full calculations

**C. Fiolhais, F. Nogueira, and M. Marques**

Chapter 6: A Tutorial on Density Functional Theory

**Journal articles on Brillouin-zone integration**

A. Baldereschi, "Mean-Value Point in the Brillouin Zone," Phys. Rev. B **7**, 5212 (1973).

D. J. Chadi, M. L. Cohen, "Special Points in the Brillouin Zone," Phys. Rev. B **8**, 5747 (1973).

H. J. Monkhorst and J. D. Pack, "Special points for Brillouin-Zone integration," Phys. Rev. B **13**, 5188 (1976).

**Journal articles on defects in semiconductors**

C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti, and C. G. Van de Walle, "First-principles calculations for point defects in solids," Rev. Mod. Phys. **86**, 253 (2014).

***Advanced Calculations for Defects in Materials***

**A. Alkauskas, P. Deák, J. Neugebauer, A. Pasquarello,**

**and C. G. Van de Walle (Eds.)**

John Wiley & Sons, 2011.

### A.3.6 VASP

**Journal articles on defects in Semiconductors**

G. Kresse and J. Furthmuller,"Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," Phys. Rev. B **54**, 169 (1996).

G. Kresse and J. Furthmller,"Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set", Comp. Mat. Sci. **6**, 15 (1996).

**Online resources**

Online manual: `http://cms.mpi.univie.ac.at/vasp/vasp/vasp.html`

Wiki: `http://cms.mpi.univie.ac.at/wiki/index.php/The_VASP_Manual`

# Bibliography


[1] O. Madelung. *Semiconductors: Data Handbook*. Springer Berlin Heidelberg, 2004.

[2] L.I. Berger. *Semiconductor Materials*. Taylor & Francis, 1996.

[3] T. Akasaka and T. Makimoto. Flow-rate modulation epitaxy of wurtzite AlBN. *Appl. Phys. Lett.*, 88(4):041902, 2006.

[4] J.F. Nye. *Physical properties of crystals: their representation by tensors and matrices*. Clarendon Press, 1972.

[5] I. Vurgaftman and J. R. Meyer. Band parameters for nitrogen-containing semiconductors. *J. Appl. Phys.*, 94(6):3675, 2003.

[6] H. Schulz and K.H. Thiemann. Crystal structure refinement of AlN and GaN. *Solid State Commun.*, 23(11):815, 1977.

[7] F. P. Bundy and R. H. Wentorf. Direct Transformation of Hexagonal Boron Nitride to Denser Forms. *J. Chem. Phys.*, 38(5):1144, 1963.

[8] T. Soma, A. Sawaoka, and S. Saito. Characterization of wurtzite type boron nitride synthesized by shock compression. *Mater. Res. Bull.*, 9(6):755, 1974.

[9] A. Nagakubo, H. Ogi, H. Sumiya, K. Kusakabe, and M. Hirao. Elastic constants of cubic and wurtzite boron nitrides. *Appl. Phys. Lett.*, 102(24):241909, 2013.

[10] W. Paszkowicz, J. B. Pelka, and M. Knapp. Lattice parameters and anisotropic thermal expansion of hexagonal boron nitride in the 10–297.5 K temperature range. *Applied Physics A*, 75(3):431, 2002.

[11] D. A. Evans, A. G. McGlynn, B. M. Towlson, M. Gunn, D. Jones, T. E. Jenkins, R. Winter, and N. R. J. Poolton. Determination of the optical band-gap energy of cubic and hexagonal boron nitride using luminescence excitation spectroscopy. *J. Phys: Condens. Mat.*, 20(7):075233, 2008.

[12] K. Lawniczak-Jablonska, T. Suski, I. Gorczyca, N. E. Christensen, K. E. Attenkofer, R. C. C. Perera, E. M. Gullikson, J. H. Underwood, D. L. Ederer, and Z. Liliental Weber. Electronic states in valence and conduction bands of group-III nitrides: Experiment and theory. *Phys. Rev. B*, 61(24):16623, 2000.

[13] K. Watanabe, T. Taniguchi, and H. Kanda. Direct-bandgap properties and evidence for ultraviolet lasing of hexagonal boron nitride single crystal. *Nat. Mater.*, 3(6):404, 2004.

[14] C. Tarrio and S. Schnatterly. Interband transitions, plasmons, and dispersion in hexagonal boron nitride. *Phys. Rev. B*, 40(11):7852, 1989.

[15] V. L. Solozhenko, A. G. Lazarenko, J.-P. Petitet, and A. V. Kanaev. Bandgap energy of graphite-like hexagonal boron nitride. *J. Phys. Chem. Solids*, 62(7):1331, 2001.

[16] L. Teles, J. Furthmüller, L. Scolfaro, J. Leite, and F Bechstedt. First-principles calculations of the thermodynamic and structural properties of strained $In_xGa_{1-x}N$ and $Al_xGa_{1-x}N$ alloys. *Phys. Rev. B*, 62(4):2475, 2000.

[17] S. Pimputkar, J. S. Speck, S. P. DenBaars, and S. Nakamura. Prospects for LED lighting. *Nat. Photonics*, 3(4):180, 2009.

[18] U.K. Mishra and P. Parikh. AlGaN/GaN HEMTs–An overview of device operation and applications. *P. IEEE*, 90(6):1022, 2002.

[19] L. S. McCarthy, P. Kozodoy, M. J. W. Rodwell, S. P. DenBaars, and U. K. Mishra. AlGaN/GaN heterojunction bipolar transistor. *IEEE Electr. Device Lett.*, 20(6):277, 1999.

[20] T. Paskova, D. A. Hanser, and K. R. Evans. GaN Substrates for III-Nitride Devices. *P. IEEE*, 98(7):1324, 2010.

[21] T. Hashimoto, F. Wu, J. S. Speck, and S. Nakamura. A GaN bulk crystal with improved structural quality grown by the ammonothermal method. *Nat. Mater.*, 6(8):568, 2007.

[22] E. V. Etzkorn and D. R. Clarke. Cracking of GaN films. *J. Appl. Phys.*, 89(2):1025, 2001.

[23] J. W. Hutchinson and Z Suo. Mixed Mode Cracking in Layered Materials. *Adv. Appl. Mech.*, 29:63, 1992.

[24] C. E. Dreyer, A. Janotti, and C. G. Van de Walle. Effects of strain on the electron effective mass in GaN and AlN. *Appl. Phys. Lett.*, 102(14):142105, 2013.

[25] H. Morkoç. *Nitride Semiconductors and Devices*. Springer Berlin Heidelberg, 1999.

[26] C. E. Dreyer, A. Janotti, and C. G. Van de Walle. Brittle fracture toughnesses of GaN and AlN from first-principle surface energy calculations. *Unpublished.*

[27] Y. Mori, Y. Kitaoka, M. Imade, N. Miyoshi, M. Yoshimura, and T. Sasaki. Growth of bulk GaN crystals by Na flux method. *Phys. Stat. Sol. c*, 8(5):1445, 2011.

[28] F. Qian, Y. Li, S. Gradecak, H.-G. Park, Y. Dong, Y. Ding, Z. L. Wang, and C. M. Lieber. Multi-quantum-well nanowire heterostructures for wavelength-controlled lasers. *Nat. Mater.*, 7(9):701, 2008.

[29] C. E. Dreyer, A. Janotti, and C. G. Van de Walle. Absolute surface energies of polar and nonpolar planes of GaN. *Phys. Rev. B*, 89(8):081305, 2014.

[30] O. Ambacher, R. Dimitrov, M. Stutzmann, B. E. Foutz, M. J. Murphy, J. A. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Chumbes, B. Green, A. J. Sierakowski, W. J. Schaff, and L. F. Eastman. Role of Spontaneous and Piezoelectric Polarization Induced Effects in Group-III Nitride Based Heterostructures and Devices. *Phys. Stat. Sol. (b)*, 216(1):381, 1999.

[31] E. Kioupakis, Q. Yan, and C. G. Van de Walle. Interplay of polarization fields and Auger recombination in the efficiency droop of nitride light-emitting diodes. *Appl. Phys. Lett.*, 101(23):231107, 2012.

[32] C. E. Dreyer, A. Janotti, and C. G. Van de Walle. Spontaneous polarization constants in III-nitrides and the importance of a consistent reference structure. *unpublished.*

[33] R.M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, 2004.

[34] P. Hohenberg and W Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864, 1964.

[35] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133, 1965.

[36] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45(7):566, 1980.

[37] J. P. Perdew and A. Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23(10):5048, 1981.

[38] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865, 1996.

[39] M. Levy, J. P. Perdew, and V. Sahni. Exact differential equation for the density and ionization energy of a many-particle system. *Phys. Rev. A*, 30(5):2745, 1984.

[40] W. G. Aulbur, L. Jönsson, and J. W. Wilkins. Quasiparticle calculations in solids. volume 54. Academic Press, 1999.

[41] A. D. Becke. A new mixing of Hartree-Fock and local density functional theories. *J. Chem. Phys.*, 98(2):1372, 1993.

[42] J. Harris. Adiabatic-connection approach to Kohn-Sham theory. *Phys. Rev.A*, 29(4):1648, 1984.

[43] J. Heyd, G. E. Scuseria, and M. Ernzerhof. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.*, 118(18):8207, 2003.

[44] L. Cláudio de Carvalho, A. Schleife, and F. Bechstedt. Influence of exchange and correlation on structural and electronic properties of AlN, GaN, and InN polytypes. *Phys. Rev. B*, 84(19):195105, 2011.

[45] J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Angyán. Screened hybrid density functionals applied to solids. *J. Chem. Phys.*, 124(15):154709, 2006.

[46] J. Heyd, G. E. Scuseria, and M. Ernzerhof. Erratum: Hybrid functionals based on a screened Coulomb potential [J. Chem. Phys. 118, 8207 (2003)]. *J. Chem. Phys.*, 124(21):219906, 2006.

[47] J. L. Lyons, A. Janotti, and C. G. Van de Walle. Carbon impurities and the yellow luminescence in GaN. *Appl. Phys. Lett.*, 97(15):152108, 2010.

[48] P. G. Moses, M. Miao, Q. Yan, and C. G. Van de Walle. Hybrid functional investigations of band gaps and band alignments for AlN, GaN, InN, and InGaN. *J. Chem. Phys.*, 134(8):084703, 2011.

[49] G. Kresse and J. Furthmüller. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54(16):11169, 1996.

[50] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50(24):17953, 1994.

[51] A. Baldereschi. Mean-Value Point in the Brillouin Zone. *Phys. Rev. B*, 7(12):5212, 1973.

[52] D. J. Chadi and M. Cohen. Special Points in the Brillouin Zone. *Phys. Rev. B*, 8(12):5747, 1973.

[53] H. J. Monkhorst and J. D. Pack. Special points for Brillouin-zone integrations. *Phys. Rev. B*, 13(12):5188, 1976.

[54] S. E. Thompson, S. Suthram, Y. Sun, G. Sun, S. Parthasarathy, M. Chu, and T. Nishida. Future of Strained Si/Semiconductors in Nanoscale MOSFETs. In *2006 International Electron Devices Meeting*, page 1. IEEE, 2006.

[55] M. Azize and T. Palacios. Effect of substrate-induced strain in the transport properties of AlGaN/GaN heterostructures. *J. Appl. Phys.*, 108(2):023707, 2010.

[56] Q. Yan, P. Rinke, M. Scheffler, and C. G. Van de Walle. Strain effects in group-III nitrides: Deformation potentials for AlN, GaN, and InN. *Appl. Phys. Lett.*, 95(12):121111, 2009.

[57] H. Kroemer. *Quantum mechanics for engineering, materials science, and applied physics*. Prentice Hall, 1994.

[58] E. O. Kane. Band structure of indium antimonide. *J. Phys. Chem. Solids*, 1(4):249, 1957.

[59] S. Chuang and C. Chang. k·p method for strained wurtzite semiconductors. *Phys. Rev. B*, 54(4):2491, 1996.

[60] G. L. Bir and G. E. Pikus. *Symmetry and Strain-induced Effects in Semiconductors*. Wiley/Halsted Press, 1974.

[61] I. Gorczyca, J. Plesiewicz, L. Dmowski, T. Suski, N. E. Christensen, A. Svane, C. S. Gallinat, G. Koblmueller, and J. S. Speck. Electronic structure and effective masses of InN under pressure. *J. Appl. Phys.*, 104(1):013704, 2008.

[62] A. Svane, N. E. Christensen, I. Gorczyca, M. van Schilfgaarde, A. N. Chantis, and T. Kotani. Quasiparticle self-consistent GW theory of III-V nitride semiconductors: Bands, gap bowing, and effective masses. *Phys. Rev. B*, 82(11):115102, 2010.

[63] P. Rinke, M. Winkelnkemper, A. Qteish, D. Bimberg, J. Neugebauer, and M. Scheffler. Consistent set of band parameters for the group-III nitrides AlN, GaN, and InN. *Phys. Rev. B*, 77(7):075202, 2008.

[64] M. Ueno, M. Yoshida, A. Onodera, O. Shimomura, and K. Takemura. Stability of the wurtzite-type structure under high pressure: GaN and InN. *Phys. Rev. B*, 49(1):14, 1994.

[65] G. D. Chen, M. Smith, J. Y. Lin, H. X. Jiang, S.-H. Wei, M. Asif Khan, and C. J. Sun. Fundamental optical transitions in GaN. *Appl. Phys. Lett.*, 68(20):2784, 1996.

[66] A. Shikanai, T. Azuhata, T. Sota, S. F. Chichibu, A. Kuramata, K. Horino, and S. Nakamura. Biaxial strain dependence of exciton resonance energies in wurtzite GaN. *J. Appl. Phys.*, 81(1):417, 1997.

[67] H. Ikeda, T. Okamura, K. Matsukawa, T. Sota, M. Sugawara, T. Hoshi, P. Cantu, R. Sharma, J. F. Kaeding, S. Keller, U. K. Mishra, K. Kosaka, K. Asai, S. Sumiya, T. Shibata, M. Tanaka, J. S. Speck, S. P. DenBaars, S. Nakamura, T. Koyama, T. Onuma, and S. F. Chichibu. Impact of strain on free-exciton resonance energies in wurtzite AlN. *J. Appl. Phys.*, 102(12):123707, 2007.

[68] N.W. Ashcroft and N.D. Mermin. *Solid state physics*. Saunders College, 1976.

[69] C. Herring. Some Theorems on the Free Energies of Crystal Surfaces. *Phys. Rev.*, 82(1):87, 1951.

[70] P. Hartman and W. G. Perdok. On the relations between structure and morphology of crystals I. *Acta Crystallogr.*, 8(1):49, 1955.

[71] Q. Sun, C. D. Yerino, B. Leung, J. Han, and M. E. Coltrin. Understanding and controlling heteroepitaxy with the kinetic Wulff plot: A case study with GaN. *J. Appl. Phys.*, 110(5):053517, 2011.

[72] B. N. Bryant, A. Hirai, E. C. Young, S. Nakamura, and J. S. Speck. Quasi-equilibrium crystal shapes and kinetic Wulff plots for gallium nitride grown by hydride vapor phase epitaxy. *J. Cryst. Growth*, 369:14, 2013.

[73] S. Pimputkar, S. Kawabata, J. S. Speck, and S. Nakamura. Surface morphology study of basic ammonothermal GaN grown on non-polar GaN seed crystals of varying surface orientations from m-plane to a-plane. *J. Cryst. Growth*, 368:67, 2013.

[74] E. Arbel and J. W. Cahn. On invariances in surface thermodynamic properties and their applications to low symmetry crystals. *Surf. Sci.*, 51(1):305, 1975.

[75] N. Chetty and R. M. Martin. Determination of integrals at surfaces using the bulk crystal symmetry. *Phys. Rev. B*, 44(11):5568, 1991.

[76] S. Zhang and S.-H. Wei. Surface Energy and the Common Dangling Bond Rule for Semiconductors. *Phys. Rev. Lett.*, 92(8):086102, 2004.

[77] L. Manna, L. W. Wang, R. Cingolani, and A. P. Alivisatos. First-principles modeling of unpassivated and surfactant-passivated bulk facets of wurtzite CdSe: A model system for studying the anisotropic growth of CdSe nanocrystals. *J. Phys. Chem. B*, 109(13):6183, 2005.

[78] J. Y. Rempel, B. L. Trout, M. G. Bawendi, and K. F. Jensen. Properties of the CdSe(0001), $(000\bar{1})$, and $(11\bar{2}0)$ single crystal surfaces: Relaxation, reconstruction, and adatom and admolecule adsorption. *J. Phys. Chem. B*, 109(41):19320, 2005.

[79] D. Segev and C. G. Van de Walle. Origins of Fermi-level pinning on GaN and InN polar and nonpolar surfaces. *Europhys. Lett.*, 76(2):305, 2006.

[80] D. Segev and C. G. Van de Walle. Electronic structure of nitride surfaces. *J. Cryst. Growth*, 300(1):199, 2007.

[81] A. R. Smith, R. M. Feenstra, D. W. Greve, J. Neugebauer, and J. E. Northrup. Reconstructions of the GaN$(000\bar{1})$ Surface. *Phys. Rev. Lett.*, 79(20):3934, 1997.

[82] A. R. Smith, R. M. Feenstra, D. W. Greve, M.-S. Shin, M. Skowronski, J. Neugebauer, and J. E. Northrup. Reconstructions of GaN (0001) and $(000\bar{1})$ surfaces: Ga-rich metallic structures. *J. Vac. Sci. Technol. B*, 16(4):2242, 1998.

[83] A. R. Smith, R. M. Feenstra, D. W. Greve, M.-S. Shin, M. Skowronski, J. Neugebauer, and J.E. Northrup. GaN(0001) surface structures studied using scanning tunneling microscopy and first-principles total energy calculations. *Surf. Sci.*, 423(1):70, 1999.

[84] David Segev and C. G. Van de Walle. Surface reconstructions on InN and GaN polar and nonpolar surfaces. *Surf. Sci.*, 601(4):L15, 2007.

[85] J. E. Northrup and J. Neugebauer. Indium-induced changes in GaN(0001) surface morphology. *Phys. Rev. B*, 60(12):R8473, 1999.

[86] J. E. Northrup, R. Di Felice, and J. Neugebauer. Energetics of H and $NH_2$ on GaN($10\bar{1}0$) and implications for the origin of nanopipe defects. *Phys. Rev. B*, 56(8):R4325, 1997.

[87] C. G. Van de Walle and J. Neugebauer. First-principles surface phase diagram for hydrogen on GaN surfaces. *Phys. Rev. Lett.*, 88(6):066103, 2002.

[88] J. E. Northrup and J. Neugebauer. Strong affinity of hydrogen for the GaN($000\bar{1}$) surface: Implications for molecular beam epitaxy and metalorganic chemical vapor deposition. *Appl. Phys. Lett.*, 85(16):3429, 2004.

[89] V. Jindal and F. Shahedipour-Sandvik. Theoretical prediction of GaN nanostructure equilibrium and nonequilibrium shapes. *J. Appl. Phys.*, 106(8):083115, 2009.

[90] B. Leung, Q. Sun, C. D. Yerino, J. Han, and M. E. Coltrin. Using the kinetic Wulff plot to design and control nonpolar and semipolar GaN heteroepitaxy. *Semicond. Sci. Tech.*, 27(2):024005, 2012.

[91] D. Du, D. J. Srolovitz, M. E. Coltrin, and C. C. Mitchell. Systematic prediction of kinetically limited crystal growth morphologies. *Phys. Rev. Lett.*, 95(15):155503, 2005.

[92] E. C. Young, A. E. Romanov, C. S. Gallinat, A. Hirai, G. E. Beltz, and J. S. Speck. Anisotropy of tensile stresses and cracking in nonbasal plane $Al_xGa_{1-x}N$/GaN heterostructures. *Appl. Phys. Lett.*, 96(4):041913, 2010.

[93] A. E. Romanov, E. C. Young, F. Wu, A. Tyagi, C. S. Gallinat, S. Nakamura, S. P. DenBaars, and J. S. Speck. Basal plane misfit dislocations and stress relaxation in III-nitride semipolar heteroepitaxy. *J. Appl. Phys.*, 109(10):103522, 2011.

[94] J. S. Speck and S. F. Chichibu. Nonpolar and semipolar group III nitride-based materials. *MRS Bull.*, 34(05):304, 2009.

[95] J. W. Matthews and A. E. Blakeslee. Defects in epitaxial multilayers. *J. Cryst. Growth*, 27:118, 1974.

[96] D. Chidambarrao, G. R. Srinivasan, B. Cunningham, and C. S. Murthy. Effects of Peierls barrier and epithreading dislocation orientation on the critical thickness in heteroepitaxial structures. *Appl. Phys. Lett.*, 57(10):1001, 1990.

[97] S. Yoshida, T. Yokogawa, Y. Imai, S. Kimura, and O. Sakata. Evidence of lattice tilt and slip in *m*-plane InGaN/GaN heterostructure. *Appl. Phys. Lett.*, 99(13):131909, 2011.

[98] B. Jahnen, M. Albrecht, W. Dorsch, S. Christiansen, and H. P. Strunk. Pinholes, dislocations and strain relaxation in InGaN. *MRS Internet J. Nitride Semicond. Res.*, 3:39, 1998.

[99] J. A. Floro, D. M. Follstaedt, P. Provencio, S. J. Hearne, and S. R. Lee. Misfit dislocation formation in the AlGaN/GaN heterointerface. *J. Appl. Phys.*, 96(12):7087, 2004.

[100] S. J. Hearne, J. Han, S. R. Lee, J. A. Floro, D. M. Follstaedt, E. Chason, and I. S. T. Tsong. Brittle-ductile relaxation kinetics of strained AlGaN/GaN heterostructures. *Appl. Phys. Lett.*, 76(12):1534, 2000.

[101] S. Einfeldt, V. Kirchner, H. Heinke, M. Dießelberg, S. Figge, K. Vogeler, and D. Hommel. Strain relaxation in AlGaN under tensile plane stress. *J. Appl. Phys.*, 88(12):7029, 2000.

[102] J.-M. Bethoux, P. Venngus, F. Natali, E. Feltin, O. Tottereau, G. Nataf, P. De Mierry, and F. Semond. Growth of high quality crack-free AlGaN films on GaN templates using plastic relaxation through buried cracks. *J. Appl. Phys.*, 94(10):6499, 2003.

[103] S. R. Lee, D. D. Koleske, K. C. Cross, J. A. Floro, K. E. Waldrip, A. T. Wise, and S. Mahajan. *In situ* measurements of the critical thickness for strain relaxation in AlGaN/GaN heterostructures. *Appl. Phys. Lett.*, 85(25):6164, 2004.

[104] K. Ito, K. Hiramatsu, H. Amano, and I. Akasaki. Preparation of $Al_xGa_{1-x}N$/GaN heterostructure by MOVPE. *J. Cryst. Growth*, 104(2):533, 1990.

[105] J. Qu, J. Li, and G. Zhang. AlGaN/GaN heterostructure grown by metalorganic vapor phase epitaxy. *Solid State Commun.*, 107(9):467, 1998.

[106] A. Bourret, C. Adelmann, B. Daudin, J.-L. Rouvière, G. Feuillet, and G. Mula. Strain relaxation in (0001) AlN/GaN heterostructures. *Phys. Rev. B*, 63(24):245307, 2001.

[107] P. J. Parbrook, T. Wang, M. A. Whitehead, C. N. Harrison, R. J. Lynch, and R. T. Murray. Crack formation and development in AlGaN/GaN structures. *Phys. Stat. Sol. (c)*, 0(7):2055, 2003.

[108] S. Terao, M. Iwaya, R. Nakamura, S. Kamiyama, H. Amano, and I. Akasaki. Fracture of $Al_xGa_{1-x}N$/GaN heterostructure – Compositional and impurity dependence. *Jpn. J. Appl. Phys.*, 40(3A):L195, 2001.

[109] M. Miao, A. Janotti, and C. G. Van de Walle. Reconstructions and origin of surface states on AlN polar and nonpolar surfaces. *Phys. Rev. B*, 80(15):155319, 2009.

[110] J. E. Northrup and J. Neugebauer. Theory of GaN($10\bar{1}0$) and ($11\bar{2}0$) surfaces. *Phys. Rev. B*, 53(16):R10477, 1996.

[111] L.B. Freund. *Dynamic Fracture Mechanics*. Cambridge University Press, 1998.

[112] D.J. Griffiths. *Introduction to Electrodynamics*. Prentice Hall, 1999.

[113] G. Makov and M. C. Payne. Periodic boundary conditions in *ab initio* calculations. *Phys. Rev. B*, 51(7):4014, 1995.

[114] R. Resta. Theory of the electric polarization in crystals. *Ferroelectrics*, 136(1-4):51, 1992. 2nd Williamsburg Workshop on First Principles Calculations for Ferroelectrics, Williamsburg, VA Feb. 02-04, 1992.

[115] R. D. King-Smith and D. Vanderbilt. Theory of polarization of crystalline solids. *Phys. Rev. B*, 47(3):1651, 1993.

[116] K. M. Rabe, C. H. Ahn, and J. M. Triscone. *Physics of Ferroelectrics: A Modern Perspective*. Springer, 2010.

[117] R. M. Martin. Comment on calculations of electric polarization in crystals. *Phys. Rev. B*, 9(4):1998, 1974.

[118] D. Vanderbilt and R. D. King-Smith. Electronic polarization in the ultrasoft pseudopotential formalism. 1998. http://arxiv.org/abs/cond-mat/9801177.

[119] J. Zak. Berry's phase for energy bands in solids. *Phys. Rev. Lett.*, 62(23):2747, 1989.

[120] D. Vanderbilt and R. D. King-Smith. Electric polarization as a bulk quantity and its relation to surface charge. *Phys. Rev. B*, 48(7):4442, 1993.

[121] R. M. Martin and G. Ortiz. Functional theory of extended Coulomb systems. *Phys. Rev. B*, 56(3):1124, 1997.

[122] R. M. Martin. Piezoelectricity. *Phys. Rev. B*, 5(4):1607, 1972.

[123] M. Feneberg and K. Thonke. Polarization fields of III-nitrides grown in different crystal orientations. *Journal of physics. Condensed matter : an Institute of Physics journal*, 19(40):403201, 2007.

[124] D. Vanderbilt. Berry-phase theory of proper piezoelectric response. *J. Phys. Chem. Solids*, 61(2):147, 2000.

[125] F. Bernardini, V. Fiorentini, and D. Vanderbilt. Spontaneous polarization and piezoelectric constants of III-V nitrides. *Phys. Rev. B*, 56(16):R10024, 1997.

[126] F. Bernardini, V. Fiorentini, and D. Vanderbilt. Accurate calculation of polarization-related quantities in semiconductors. *Phys. Rev. B*, 63(19):193201, 2001.

[127] F. Bechstedt, U. Grossner, and J. Furthmüller. Dynamics and polarization of group-III nitride lattices: A first-principles study. *Phys. Rev. B*, 62(12):8003, 2000.

[128] J. W. Bennett, K. F. Garrity, K. M. Rabe, and D. Vanderbilt. Hexagonal ABC Semiconductors as Ferroelectrics. *Phys. Rev. Lett.*, 109(16):167602, 2012.

[129] E. T. Yu, G. J. Sullivan, P. M. Asbeck, C. D. Wang, D. Qiao, and S. S. Lau. Measurement of piezoelectrically induced charge in GaN/AlGaN heterostructure field-effect transistors. *Appl. Phys. Lett.*, 71(19):2794, 1997.

[130] O. Ambacher, B. Foutz, J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murphy, A. J. Sierakowski, W. J. Schaff, L. F. Eastman, R. Dimitrov, A. Mitchell, and M. Stutzmann. Two dimensional electron gases induced by spontaneous and piezoelectric polarization in undoped and doped AlGaN/GaN heterostructures. *J. Appl. Phys.*, 87(1):334, 2000.

[131] G. Martínez-Criado, A. Cros, A. Cantarero, O. Ambacher, C. R. Miskys, R. Dimitrov, M. Stutzmann, J. Smart, and J. R. Shealy. Residual strain effects on the two-dimensional electron gas concentration of AlGaN/GaN heterostructures. *J. Appl. Phys.*, 90(9):4735, 2001.

[132] S. C. Binari, J. M. Redwing, G. Kelner, and W. Kruppa. AlGaN/GaN HEMTs grown on SiC substrates. *Electron. Lett.*, 33(3):242, 1997.

[133] Y.-F. Wu, B. P. Keller, S. Keller, D. Kapolnek, P. Kozodoy, S. P. Denbaars, and U. K. Mishra. Very high breakdown voltage and large transconductance realized on GaN heterojunction field effect transistors. *Appl. Phys. Lett.*, 69(10):1438, 1996.

[134] C. Nguyen, N. X. Nguyen, M. Le, and D. E. Grider. High performance GaN/AlGaN MODFETs grown by RF-assisted MBE. *Electron. Lett.*, 34(3):309, 1998.

[135] L. W. Wong, S. J. Cai, R. Li, K. Wang, H. W. Jiang, and M. Chen. Magnetotransport study on the two-dimensional electron gas in AlGaN/GaN heterostructures. *Appl. Phys. Lett.*, 73(10):1391, 1998.

[136] K. Köhler, S. Müller, P. Waltereit, W. Pletschen, V. Polyakov, T. Lim, L. Kirste, H. P. Menner, P. Brückner, O. Ambacher, C. Buchheim, and R. Goldhahn. Electrical properties of $Al_xGa_{1-x}N$/GaN heterostructures with low Al content. *J. Appl. Phys.*, 109(5):053705, 2011.

[137] A. Nakajima, Y. Sumida, M. H. Dhyani, H. Kawai, and E. M. S. Narayanan. High Density Two-Dimensional Hole Gas Induced by Negative Polarization at GaN/AlGaN Heterointerface. *Appl. Phys. Express*, 3(12):121004, 2010.

[138] J. Burm, W. J. Schaff, L. F. Eastman, H. Amano, and I. Akasaki. 75 Å GaN channel modulation doped field effect transistors. *Appl. Phys. Lett.*, 68(20):2849, 1996.

[139] T. Wang, Y. Ohno, M. Lachab, D. Nakagawa, T. Shirahama, S. Sakai, and H. Ohno. Electron mobility exceeding $10^4$ $cm^2/V$ s in an AlGaN/GaN heterostructure grown on a sapphire substrate. *Appl. Phys. Lett.*, 74(23):3531, 1999.

[140] Y. Liu, M. Z. Kauser, D. D. Schroepfer, P. P. Ruden, J. Xie, Y. T. Moon, N. Onojima, H. Morkoç, K.-A. Son, and M. I. Nathan. Effect of hydrostatic pressure on the current-voltage characteristics of GaN/AlGaN/GaN heterostructure devices. *J. Appl. Phys.*, 99(11):113706, 2006.

[141] G. J. Ding, L. W. Guo, Z. G. Xing, Y. Chen, P. Q. Xu, H. Q. Jia, J. M. Zhou, and H. Chen. Characterization of different-Al-content AlGaN/GaN heterostructures on sapphire. *Sci. China Phys. Mech. Astron.*, 53(1):49–, 2010.

[142] W. S. Chen, S. J. Chang, Y. K. Su, R. L. Wang, C. H. Kuo, and S. C. Shei. $Al_xGa_{1-x}N$/GaN heterostructure field effect transistors with various Al mole fractions in AlGaN barrier. *J. Cryst. Growth*, 275(3-4):398, 2005.

[143] K. Jeganathan, T. Ide, M. Shimizu, and H. Okumura. Two-dimensional electron gases induced by polarization charges in AlN/GaN heterostructure grown by plasma-assisted molecular-beam epitaxy. *J. Appl. Phys.*, 94(5):3260, 2003.

[144] R. Gaska, A. Osinsky, J. W. Yang, and M. S. Shur. Self-Heating in High-Power AlGaN-GaN HFETs. *IEEE Electr. Device Lett.*, 19(3):89, 1998.

[145] P. M. Asbeck, E. T. Yu, S. S. Lau, G. J. Sullivan, J. Van Hove, and J. Redwing. Piezoelectric charge densities in AlGaN/GaN HFETs. *Electron. Lett.*, 33(14):1230, 1997.

[146] I. P. Smorchkova, C. R. Elsass, J. P. Ibbetson, R. Vetury, B. Heying, P. Fini, E. Haus, S. P. DenBaars, J. S. Speck, and U. K. Mishra. Polarization-induced charge and electron mobility in AlGaN/GaN heterostructures grown by plasma-assisted molecular-beam epitaxy. *J. Appl. Phys.*, 86(8):4520, 1999.

[147] E. J. Miller, E. T. Yu, C. Poblenz, C. Elsass, and J. S. Speck. Direct measurement of the polarization charge in AlGaN/GaN heterostructures using capacitancevoltage carrier profiling. *Appl. Phys. Lett.*, 80(19):3551, 2002.

[148] L. Jia, E. T. Yu, D. Keogh, P. M. Asbeck, P. Miraglia, A. Roskowski, and R. F. Davis. Polarization charges and polarization-induced barriers in $Al_xGa_{1-x}N$/GaN and $In_yGa_{1-y}N$/GaN heterostructures. *Appl. Phys. Lett.*, 79(18):2916, 2001.

[149] G. Franssen, J. A. Plesiewicz, L. H. Dmowski, P. Prystawko, T. Suski, W. Krupczyński, R. Jachymek, P. Perlin, and M. Leszczyński. Hydrostatic pressure dependence of polarization-induced interface charge in AlGaN/GaN heterostructures determined by means of capacitance-voltage characterization. *J. Appl. Phys.*, 100(11):113712, 2006.

[150] S. K. Davidsson, M. Gurusinghe, T. G. Andersson, and H. Zirath. The influence of composition and unintentional doping on the two-dimensional electron gas density in AlGaN/GaN heterostructures. *J. of Electron. Mater.*, 33(5):440, 2004.

[151] J. Simon, R. Langer, A. Barski, M. Zervos, and N. T. Pelekanos. Residual doping effects on the amplitude of polarization-induced electric fields in GaN/AlGaN quantum wells. *Phys. Stat. Sol. (a)*, 188(2):867, 2001.

[152] N. Grandjean, B. Damilano, S. Dalmasso, M. Leroux, M. Lagt, and J. Massies. Built-in electric-field effects in wurtzite AlGaN/GaN quantum wells. *J. Appl. Phys.*, 86(7):3714, 1999.

[153] R. Cingolani, A. Botchkarev, H. Tang, H. Morkoç, G. Traetta, G. Coli, M. Lomascolo, A. Di Carlo, F. Della Sala, and P Lugli. Spontaneous polarization and piezoelectric field in $GaN/Al_{0.15}Ga_{0.85}N$ quantum wells: Impact on the optical spectra. *Phys. Rev. B*, 61(4):2711, 2000.

[154] G. Chen, Z. L. Li, X. Q. Wang, C. C. Huang, X. Rong, L. W. Sang, F. J. Xu, N. Tang, Z. X. Qin, M. Sumiya, Y. H. Chen, W. K. Ge, and B. Shen. Effect of

polarization on intersubband transition in AlGaN/GaN multiple quantum wells. *Appl. Phys. Lett.*, 102(19):192109, 2013.

[155] C. McAleese, P. M. F. J. Costa, D. M. Graham, H. Xiu, J. S. Barnard, M. J. Kappers, P. Dawson, M. J. Godfrey, and C. J. Humphreys. Electric fields in AlGaN/GaN quantum well structures. *Phys. Stat. Sol. (b)*, 243(7):1551, 2006.

[156] R. Langer, J. Simon, V. Ortiz, N. T. Pelekanos, A. Barski, R. André, and M. Godlewski. Giant electric fields in unstrained GaN single quantum wells. *Appl. Phys. Lett.*, 74(25):3827, 1999.

[157] M. Leroux, N. Grandjean, M. Laügt, J. Massies, B. Gil, P. Lefebvre, and P. Bigenwald. Quantum confined Stark effect due to built-in internal polarization fields in (Al,Ga)N/GaN quantum wells. *Phys. Rev. B*, 58(20):R13371, 1998.

[158] N. Suzuki and N. Iizuka. Effect of Polarization Field on Intersubband Transition in AlGaN/GaN Quantum Wells. *Jpn. J. Appl. Phys.*, 38(Part 2, No. 4A):L363, 1999.

[159] M. Esmaeili, M. Sabooni, H. Haratizadeh, P. P. Paskov, B. Monemar, P. O. Holz, S. Kamiyama, and M. Iwaya. Optical properties of GaN/AlGaN QW nanostructures with different well and barrier widths. *J. Phys: Condens. Mat.*, 19(35):356218, 2007.

[160] M. Esmaeili, M. Gholami, H. Haratizadeh, B. Monemar, P. O. Holtz, S. Kamiyama, H. Amano, and I. Akasaki. Experimental and theoretical investigations of optical properties of GaN/AlGaN MQW nanostructures. Impact of built-in polarization fields. *Opto-Electron. Rev.*, 17(4):293, 2009.

[161] A. Pinos, S. Marcinkevičius, K. Liu, M. S. Shur, E. Kuokštis, G. Tamulaitis, R. Gaska, J. Yang, and W. Sun. Screening dynamics of intrinsic electric field in AlGaN quantum wells. *Appl. Phys. Lett.*, 92(6):061907, 2008.

[162] S. Marcinkevičius, A. Pinos, K. Liu, D. Veksler, M. S. Shur, J. Zhang, and R. Gaska. Intrinsic electric fields in AlGaN quantum wells. *Appl. Phys. Lett.*, 90(8):081914, 2007.

[163] E. Kuokstis, C. Q. Chen, M. E. Gaevski, W. H. Sun, J. W. Yang, G. Simin, M. Asif Khan, H. P. Maruska, D. W. Hill, M. C. Chou, J. J. Gallagher, and B. Chai. Polarization effects in photoluminescence of *c*- and *m*-plane GaN/AlGaN multiple quantum wells. *Appl. Phys. Lett.*, 81(22):4130, 2002.

[164] H. M. Ng, R. Harel, S. N. G. Chu, and A. Y. Cho. The effect of built-in electric field in GaN/AlGaN quantum wells with high AIN mole fraction. *J. Electron. Mater.*, 30(3):134, 2001.

[165] J. S. Im, H. Kollmer, J. Off, A. Sohmer, F. Scholz, and A. Hangleiter. Reduction of oscillator strength due to piezoelectric fields in GaN/$Al_xGa_{1-x}N$ quantum wells. *Phys. Rev. B*, 57(16):R9435, 1998.

[166] O. Gfrörer, T. Schlüsener, V. Härle, F. Scholz, and A. Hangleiter. Mechanisms of Strain Reduction in GaN and AlGaN/GaN Epitaxial Layers. *MRS Proc.*, 449:429, 2011.

[167] F. Renner, P. Kiesel, G. H. Döhler, M. Kneissl, C. G. Van de Walle, and N. M. Johnson. Quantitative Analysis of Absorption and Field-Induced Absorption Changes in InGaN/GaN Quantum Wells. *Phys. Stat. Sol. (b)*, 234(3):742, 2002.

[168] D.-P. Wang, C.-C. Wu, and C.-C. Wu. Determination of polarization charge density on interface of AlGaN/GaN heterostructure by electroreflectance. *Appl. Phys. Lett.*, 89(16):161903, 2006.

[169] A. T. Winzer, R. Goldhahn, C. Buchheim, O. Ambacher, A. Link, M. Stutzmann, Y. Smorchkova, U. K. Mishra, and J. S. Speck. Photoreflectance studies of N- and Ga-face AlGaN/GaN heterostructures confining a polarization induced 2DEG. *Phys. Stat. Sol. (b)*, 240(2):380, 2003.

[170] S. R. Kurtz, A. A. Allerman, D. D. Koleske, A. G. Baca, and R. D. Briggs. Electronic properties of the AlGaN/GaN heterostructure and two-dimensional electron gas observed by electroreflectance. *J. Appl. Phys.*, 95(4):1888, 2004.

[171] A Hangleiter. Optical properties and polarization fields in the nitrides. *J. Lumin.*, 87-89:130, 2000.

[172] A. Barker and M. Ilegems. Infrared lattice vibrations and free-electron dispersion in GaN. *Phys. Rev. B*, 7(2):743, 1973.

[173] S. Arulkumaran, T. Egawa, H. Ishikawa, and T. Jimbo. Characterization of different-Al-content $Al_xGa_{1-x}N$/GaN heterostructures and high-electron-mobility transistors on sapphire. *J. Vac. Sci. Technol. B*, 21(2):888, 2003.

[174] K. A. Mkhoyan. Formation of a quasi-two-dimensional electron gas in GaN/$Al_xGa_{1-x}N$ heterostructures with diffuse interfaces. *J. Appl. Phys.*, 95(4):1843, 2004.

[175] Y.-G. Zhou, B. Shen, T. Someya, H.-Q. Yu, J. Liu, H.-M. Zhou, R. Zhang, Y. Shi, Y.-D. Zheng, and Y. Arakawa. Investigation of the polarization-induced charges in modulation-doped $Al_xGa_{1-x}N$/GaN heterostructures through capacitancevoltage profiling and simulation. *Jpn. J. Appl. Phys.*, 41(Part 1, No. 4B):2531, 2002.

[176] M. E. Aumer, S. F. LeBoeuf, B. F. Moody, and S. M. Bedair. Strain-induced piezoelectric field effects on light emission energy and intensity from AlInGaN/InGaN quantum wells. *Appl. Phys. Lett.*, 79(23):3803, 2001.

[177] Q. Li, S. J. Xu, M. H. Xie, S. Y. Tong, X. H. Zhang, W. Liu, and S. J. Chua. Strong screening effect of photo-generated carriers on piezoelectric field in $In_{0.13}Ga_{0.87}N/In_{0.03}Ga_{0.97}N$ quantum wells. *Jpn. J. Appl. Phys.*, 41(Part 2, No. 10A):L1093, 2002.

[178] Y. D. Jho, J. S. Yahng, E. Oh, and D. S. Kim. Measurement of piezoelectric field and tunneling times in strongly biased InGaN/GaN quantum wells. *Appl. Phys. Lett.*, 79(8):1130, 2001.

[179] A. Hangleiter, F. Hitzel, S. Lahmann, and U. Rossow. Composition dependence of polarization fields in GaInN/GaN quantum wells. *Appl. Phys. Lett.*, 83(6):1169, 2003.

[180] S. F. Chichibu, A. C. Abare, M. S. Minsky, S. Keller, S. B. Fleischer, J. E. Bowers, E. Hu, U. K. Mishra, L. A. Coldren, S. P. DenBaars, and T. Sota. Effective band gap inhomogeneity and piezoelectric field in InGaN/GaN multiquantum well structures. *Appl. Phys. Lett.*, 73(14):2006, 1998.

[181] P. Lefebvre, A. Morel, M. Gallart, T. Taliercio, J. Allgre, B. Gil, H. Mathieu, B. Damilano, N. Grandjean, and J. Massies. High internal electric field in a graded-width InGaN/GaN quantum well: Accurate determination by time-resolved photoluminescence spectroscopy. *Appl. Phys. Lett.*, 78(9):1252, 2001.

[182] P. Kiesel, F. Renner, M. Kneissl, C. G. Van de Walle, G. H. Döhler, and N. M. Johnson. Electroabsorption Spectroscopy – Direct Determination of the Strong Piezoelectric Field in InGaN/GaN Heterostructure Diodes. *Phys. Stat. Sol. (a)*, 188(1):131, 2001.

[183] R. J. Kaplar, S. R. Kurtz, D. D. Koleske, and A. J. Fischer. Electroreflectance studies of Stark shifts and polarization-induced electric fields in InGaN/GaN single quantum wells. *J. Appl. Phys.*, 95(9):4905, 2004.

[184] Y.-K. Noh, M.-D. Kim, and J.-E. Oh. Reduction of internal polarization fields in InGaN quantum wells by InGaN/AlGaN ultra-thin superlattice barriers with different indium composition. *J. Appl. Phys.*, 110(12):123108, 2011.

[185] S.-M. Kim, H. S. Oh, J. H. Baek, K.-H. Lee, G. Y. Jung, J.-H. Song, H.-J. Kim, B.-J. Ahn, D. Yanqun, and J.-H. Song. Effects of patterned sapphire substrates on piezoelectric field in blue-emitting InGaN multiple quantum wells. *IEEE Electr. Device Lett.*, 31(8):842, 2010.

[186] G. Franssen, P. Perlin, and T. Suski. Photocurrent spectroscopy as a tool for determining piezoelectric fields in $In_xGa_{1-x}N$/GaN multiple quantum well light emitting diodes. *Phys. Rev. B*, 69(4):045310, 2004.

[187] I. H. Brown, I. A. Pope, P. M. Smowton, P. Blood, J. D. Thomson, W. W. Chow, D. P. Bour, and M. Kneissl. Determination of the piezoelectric field in InGaN quantum wells. *Appl. Phys. Lett.*, 86(13):131108, 2005.

[188] C. Y. Lai, T. M. Hsu, W.-H. Chang, K.-U. Tseng, C.-M. Lee, C.-C. Chuo, and J.-I. Chyi. Direct measurement of piezoelectric field in $In_{0.23}Ga_{0.77}N$/GaN multiple quantum wells by electrotransmission spectroscopy. *J. Appl. Phys.*, 91(1):531, 2002.

[189] D. Turchinovich, P. Uhd Jepsen, B. Monozon, M. Koch, S. Lahmann, U. Rossow, and A. Hangleiter. Ultrafast polarization dynamics in biased quantum wells under strong femtosecond optical excitation. *Phys. Rev. B*, 68(24):241307, 2003.

[190] M. Thomsen, H. Jönen, U. Rossow, and A. Hangleiter. Effects of spontaneous polarization on GaInN/GaN quantum well structures. *J. Appl. Phys.*, 109(12):123710, 2011.

[191] S. Y. Davydov. Estimates of the spontaneous polarization and permittivities of AlN, GaN, InN, and SiC crystals. *Phys. Solid State*, 51(6):1231, 2009.

[192] Y.-R. Wu, R. Shivaraman, K.-C. Wang, and J. S. Speck. Analyzing the physical properties of InGaN multiple quantum well light emitting diodes from nano scale structure. *Appl. Phys. Lett.*, 101(8):083505, 2012.

[193] W. Chen and T. G. Andersson. Quantum-confined Stark shift for differently shaped quantum wells. *Semicond. Sci. Tech.*, 7(6):828, 1992.

[194] F. Bernardini and V. Fiorentini. Nonlinear Behavior of Spontaneous and Piezoelectric Polarization in III-V Nitride Alloys. *Phys. Stat. Sol. (a)*, 190(1):65, 2002.

[195] K. Shimada, T. Sota, K. Suzuki, and H. Okumura. First-principles study on piezoelectric constants in strained BN, AlN, and GaN. *Jpn. J. Appl. Phys.*, 37(12A):L1421, 1998.

[196] C. E. Dreyer, J. L Lyons, A. Janotti, and C. G. Van de Walle. Band alignments and polarization properties of BN polymorphs. *Appl. Phys. Express*, 7(3):031001, 2014.

[197] K. Karch and F. Bechstedt. *Ab initio* lattice dynamics of BN and AlN: Covalent versus ionic forces. *Phys. Rev. B*, 56(12):7404, 1997.